Low-Complexity Algorithms for Channel Estimation in Optimised Pilot-Assisted Wireless OFDM Systems

Eugene Golovins

Thesis Presented for the Degree of

Doctor of Philosophy

in the Department of Electrical Engineering

University of Cape Town

September 2009

As the candidate's supervisor, I have approved this dissertation for submission.

Name:	Neco Ventura
Signed:	_____
Date:	_____

Declaration

I hereby declare that: (1) the above thesis is my own unaided work, both in conception and execution, and that apart from the normal guidance of my supervisor, I have received no assistance apart from that stated below; (2) except as stated below, neither the substance or any part of the thesis has been submitted in the past, or is being, or is to be submitted for a degree at this University or at any other University.

I am now presenting the thesis for examination for the Degree of PhD in Electrical Engineering. I also grant the University free licence to reproduce the above thesis in whole or in part, for the purpose of research.

Eugene Golovins

Name

Date

Abstract

Orthogonal frequency division multiplexing (OFDM) has recently become a dominant transmission technology considered for the next generation fixed and mobile broadband wireless communication systems. OFDM has an advantage of lessening the severe effects of the frequency-selective (multipath) fading due to the band splitting into relatively flat fading subchannels, and allows for low-complexity transceiver implementation based on the fast Fourier transform algorithms. Combining OFDM modulation with multilevel frequency-domain symbol mapping (e.g., QAM) and spatial multiplexing (SM) over the multiple-input multiple-output (MIMO) channels, can theoretically achieve near Shannon capacity of the communication link. However, the high-rate and spectrum-efficient system implementation requires coherent detection at the receiving end that is possible only when accurate channel state information (CSI) is available. Since in practice, the response of the wireless channel is unknown and is subject to random variation with time, the receiver typically employs a channel estimator for CSI acquisition. The channel response information retrieved by the estimator is then used by the data detector and can also be fed back to the transmitter by means of in-band or out-of-band signalling, so the latter could adapt power loading, modulation and coding parameters according to the channel conditions. Thus, design of an accurate and robust channel estimator is a crucial requirement for reliable communication through the channel, which is selective in time and frequency. In a MIMO configuration, a separate channel estimator has to be associated with each transmit/receive antenna pair, making the estimation algorithm complexity a primary concern.

Pilot-assisted methods, relying on the insertion of reference symbols in certain frequencies and time slots, have been found attractive for identification of the doubly-selective radio channels from both the complexity and performance standpoint. In this dissertation, a family of the reduced-complexity estimators for the single and multiple-antenna OFDM systems is developed. The estimators are based on the transform-domain processing and have the same order of computational complexity, irrespective of the number of pilot subcarriers and their positioning. The common estimator structure represents a cascade of successive small-dimension filtering modules. The number of modules, as well as their order inside the cascade, is determined by the class of the estimator (one or two-dimensional) and availability of the channel statistics (correlation and signal-to-noise power ratio). For fine precision estimation in the multipath channels with statistics not known a priori, we propose recursive design of the filtering modules. Simulation results show that in the steady state, performance of the recursive estimators approaches that of their theoretical counterparts, which are optimal in the minimum mean square error (MMSE) sense. In contrast to the majority of the channel estimators developed so far, our modular-type architectures are suitable for the reconfigurable OFDM transceivers where the actual channel conditions influence the decision of what class of filtering algorithm to use, and how to allot pilot subcarrier positions in the band.

In the pilot-assisted transmissions, channel estimation and detection are performed separately from each other over the distinct subcarrier sets. The estimator output is used only to construct the detector transform, but not as the detector input. Since performance of both channel estimation and detection depends on the signal-to-noise power

ratio (SNR) at the corresponding subcarriers, there is a dilemma of the optimal power allocation between the data and the pilot symbols as these are conflicting requirements under the total transmit power constraint. The problem is exacerbated by the variety of channel estimators. Each kind of estimation algorithm is characterised by its own SNR gain, which in general can vary depending on the channel correlation. In this dissertation, we optimise pilot-data power allocation for the case of developed low-complexity one and two-dimensional MMSE channel estimators. The resultant contribution is manifested by the closed-form analytical expressions of the upper bound (suboptimal approximate value) on the optimal pilot-to-data power ratio (PDR) as a function of a number of design parameters (number of subcarriers, number of pilots, number of transmit antennas, effective order of the channel model, maximum Doppler shift, SNR, etc.). The resultant PDR equations can be applied to the MIMO-OFDM systems with arbitrary arrangement of the pilot subcarriers, operating in an arbitrary multipath fading channel. These properties and relatively simple functional representation of the derived analytical PDR expressions are designated to alleviate the challenging task of on-the-fly optimisation of the adaptive SM-MIMO-OFDM system, which is capable of adjusting transmit signal configuration (e.g., block length, number of pilot subcarriers or antennas) according to the established channel conditions.

Acknowledgements

I wish to convey sincere appreciation to a number of persons for their support in many different ways during my PhD studies.

I would like to express deep gratitude to my supervisor Neco Ventura for all the financial support and research encouragement. Without his material assistance completion of this thesis would not be possible. I wish good luck to Neco in running the Centre of Excellence for Broadband Networks and every success of the future research projects under its wings.

I am very thankful to my colleagues Gabriel Andrews and Richard Good for proofreading the manuscript of dissertation and preceding publications. I hope that doing this gave them some writing experience while working on their own theses.

With all my heart I thank my parents Nina Harding and Victor Golovin for their love and care for me in the almost daily correspondence. Staying in South Africa, dozens of thousands kilometres away from my homeland Latvia has strongly convinced me of the utmost importance of the advanced communication means for the modern society. One day a person may wake up and realise that half of his/her life is going on in an endless stream of emails, SMSes and postcards, or possibly voice-over-IP calls and video conference sessions…

A special gratitude is to be tributed to my MSc supervisor Valentin Popov for all the wise advices and strong friendship through all these years. It is him and other staff members of Riga Technical University in Latvia who, being committed teachers, introduced me to the enthralling area of communication theory and signal processing.

Finally, I would like to acknowledge all the past and present members of the Communications Research Group at the University of Cape Town. I will always remember sharing professional and cultural experiences with them and having joy together. I am especially grateful to my PhD colleague and old friend Vitalis Ozianyi, whose enthusiasm, firmness of purpose and good sense of humor have been indispensable to encourage optimism in myself.

Table of Contents

List of Figures

List of Tables

Acronyms

ACF	AutoCorrelation Function
ANSSE	Average Normalised Sum of Square Errors
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
CDMA	Code Division Multiple Access
CFR	Channel Frequency Response
CGRV	Complex Gaussian Random Variable
CIR	Channel Impulse Response
CLS	Constrained Least Squares
CMMSE	Constrained Minimum Mean Square Error
COFDM	Coded Orthogonal Frequency Division Multiplexing
CP	Cyclic Prefix
CR	Cognitive Radio
CSI	Channel State Information
DCT	Discrete Cosine Transform
DD	Decision Directed
DFT	Discrete Fourier Transform
DLL	Delay Locked Loop
DMT	Discrete MultiTone
DTFT	Discrete-Time Fourier Transform
DVB-T	Digital Video Broadcasting - Terrestrial
EVD	EigenValue Decomposition
FCF	Frequency Correlation Function
FEC	Forward Error Correction
FF	Forgetting Factor
FFR	Fractional Frequency Reuse
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
GMC	Generalised MultiCarrier
GSO	Gram-Schmidt Orthogonalisation
HF	High Frequency
IDFT	Inverse Discrete Fourier Transform
IFFT	Inverse Fast Fourier Transform
IBI	Inter-Block Interference

ICI	Inter-Carrier Interference
IID	Independent Identically Distributed
ISI	Inter-Symbol Interference
LMS	Least Mean Squares
LOS	Line-Of-Sight
LS	Least Squares
LTV	Linear Time-Varying
MA	Multiple Access
MAI	Multiple Access Interference
MC	MultiCarrier
MDL	Minimum Description Length
MIMO	Multiple-Input Multiple-Output
MIP	Multipath Intensity Profile
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
NLMS	Normalised Least Mean Squares
NLOS	Non-Line-Of-Sight
NMSE	Normalised Mean Square Error
NSSE	Normalised Sum of Square Errors
OFDM	Orthogonal Frequency Division Multiplexing
PAR	Peak-to-Average power Ratio
PHY	PHYsical layer
PDF	Probability Density Function
PDP	Power-Delay Profile
PDR	Pilot-to-Data power Ratio
PSAM	Pilot-Symbol Assisted Modulation
PSD	Power Spectral Density
QAM	Quadrature Amplitude Modulation
QPSK	Quaternary Phase Shift Keying
QRD	QR Decomposition
RF	Radio Frequency
RMMSE	Recursive Minimum Mean Square Error
RMS	Root-Mean-Square
Rx	Receive
SER	Symbol Error Rate
SISO	Single-Input Single-Output

SL	Spatial Layer
SM	Spatial Multiplexing
SINR	Signal-to-Interference-plus-Noise power Ratio
SNR	Signal-to-Noise power Ratio
STBC	Space-Time Block Code
STTC	Space-Time Trellis Code
STFCF	Spaced Time-Frequency Correlation Function
SVD	Singular Value Decomposition
TB	Transmit Beamforming
TDL	Tapped Delay Line
TDM	Time Division Multiplexing
TEQ	Time-domain EQualisation
Tx	Transmit
UHF	Ultra High Frequency
UNII	Unlicenced National Information Infrastructure
UWB	Ultra-WideBand
V-BLAST	Vertical Bell Labs Layered Architecture for Space-Time
VC	Virtual subCarrier
VS	Virtual Symbol
WGN	White Gaussian Noise
WSSUS	Wide Sense Stationary Uncorrelated Scattering
ZF	Zero-Forcing

Notations

$\in$	belongs to
$\notin$	does not belong to
$\forall$	for all
j	imaginary unit
$\Re(.)$	real part
$\Im(.)$	imaginary part
$\arg(.)$	argument
$\mathrm{E}(.)$	mathematical expectation
$\lfloor . \rfloor$	rounding to smaller integer
$\lceil . \rceil$	rounding to larger integer
$a \bmod N$	division of a modulo N
$\mathbf{a} = [a_m]_{M\times 1}$	vector consisting of elements a_m, $m \in [0, M-1]$
$\mathbf{A} = [a_{m,n}]_{M\times N}$	matrix consisting of elements $a_{m,n}$, $m \in [0, M-1]$ (row index), $n \in [0, N-1]$ (column index)
$(\mathbf{A})_n$	nth column of matrix $\mathbf{A}$
$[\mathbf{A}]_{m,n}$	(m,n)th element of matrix $\mathbf{A}$
$\mathbf{I}_{M\times N}$	identity matrix of the $M \times N$ size
$\mathbf{0}_{M\times N}$	null matrix of the $M \times N$ size
$\mathbf{J}$	exchange matrix
$\mathrm{dg}(\mathbf{a})$	diagonal matrix with the elements of $\mathbf{a}$ on the main diagonal
$\mathrm{dg}(\mathbf{A})$	vector containing the diagonal elements of $\mathbf{A}$
$\mathrm{toe}(\mathbf{a}, \mathbf{b}^{\mathrm{T}})$	Toeplitz matrix with the first column $\mathbf{a}$ and the first row $\mathbf{b}^{\mathrm{T}}$
$\mathbf{A}^{\mathrm{T}}$	transpose of $\mathbf{A}$
$\mathbf{A}^{\mathrm{H}}$	Hermitean transpose of $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of $\mathbf{A}$
$\mathbf{A}^{+}$	Moore-Penrose pseudoinverse of $\mathbf{A}$
$\mathbf{A}^{2}$	$\mathbf{A}^2 = \mathbf{A}\,\mathbf{A}$
$\mathbf{A}^{1/2}$	$\mathbf{A} = \mathbf{A}^{1/2}\mathbf{A}^{1/2}$
$\mathbf{A} \circ \mathbf{B}$	Hadamard (element-wise) product
$[\mathbf{A} \quad \mathbf{B}]$	concatenation of matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathrm{tr}(.)$	trace

$\|.\|$	norm
rk(.)	rank
∇	gradient

1 Introduction

About two decades ago, the telecommunications infrastructure was dominated by fixed analogue telephony, with support for voice and narrowband data communication. The situation changed completely, with the emergence of high-speed digital network connections such as digital subscriber line (DSL) and cable modems, in the 1990s. These technologies provided internet access data rates of several megabits per second, compared to the 56kbps rates for dial-up modems working over the analogue telephone lines. In the same time frame, digital mobile communication experienced extraordinary success: the number of subscribers doubled every two years. One of the major driving forces of this evolution was the unified GSM standard, which enabled worldwide deployment of second generation (2G) mobile cellular networks. The third generation (3G) wide area wireless networks based on the wideband code division multiple access (WCDMA) guaranteed up to 2Mbps to stationary users, while offering up to 384kbps for fast moving terminals. The latest high-speed downlink packet access (HSDPA) enhancement of the 3G air interface promises peak data rates in excess of 10Mbps. Essentially, digital modems facilitate high-speed, high-reliability, and high-quality access to networked data and information services, even when the user is fully mobile.

Tremendous consumer interest in multimedia communication applications has grown rapidly during the last decade, fuelling the need for higher data rates in wired and wireless networks. The demand for wireless communications is expected to increase in the future due to the introduction of new services, like live video transmission, mobile games, and various location-based services [1]. This poses a challenge as data rates in wireless networks are limited by the need to address wide coverage, vehicular mobility and scarcity of the radio spectrum. Thus, future wireless communication systems must provide higher bandwidth and radio resource management efficiency, and accommodate faster terminal speeds. Research toward the introduction of future wireless communication systems, including beyond third (B3G) and fourth-generation (4G) cellular systems as well as next-generation mobile broadband wireless access (BWA), has already been initiated worldwide. The goal is to enable services requiring one or two orders of magnitude higher data rates than 3G systems: from 100Mbps in high-mobility outdoor applications to 1Gbps in low-mobility indoor channels for the available frequency spectrum on the order of 100MHz. This requires significant improvements in bandwidth efficiency – up to 10bps/Hz. Furthermore, the future mobile communication systems should have an open packet-oriented network architecture and allow for a full range

of multirate services, with different quality of service (QoS) and performance requirements. 3G systems are currently unable to address these requirements due to air interface restrictions.

1.1 Multicarrier system – an efficient framework for broadband wireless communication

It is widely recognised that the most performance-challenging property of the wireless medium is frequency-selective multipath propagation. It results in a portion of transmitted signal harmonics arriving at the receiver with heavily attenuated amplitudes, due to the destructive combination of the signal replicas traversing different paths. *Multicarrier* (MC) communication [2] is a natural and promising solution aimed at solving the frequency selectivity problem. The transmission is divided into parallel subchannels whose bandwidth is narrow enough to make them (at least almost) frequency-flat. By coordinating the transmission of (coded) data streams through different subchannels, one is able to avoid spectrum notches inherent in the wideband radio channel. An intuitive example of a MC system is the well-known frequency-division multiplexing (FDM) scheme, in which user's information is transmitted jointly on the frequency carriers, which do not experience deep fades.

It is an interesting fact that MC transmission, as the optimum method to solve complex transmission problems, was identified already in early Shannon's works of 1948. An early commercial system that employed this concept was Kineplex designed in 1957 for HF communication, which used four-phase differential modulation for parallel data transmission over 20 subchannels within the voice-band, and achieved a data rate of 3kbps. This rate was 10 times faster than that achieved by other single-carrier commercial modems for over 10 years.

1.1.1 Overview of multicarrier applications and standards

In general, spectral partitioning for MC transmissions can be realised in the form of overlapping or non-overlapping carrier subbands. Clearly, the receiver must be able to separate subbands before performing data demodulation. The earliest MC modems operated according to conventional FDM principle and used analogue passband filters for complete subband decoupling. Because of the difficulty of implementing filters with very sharp cutoff, such systems required the introduction of substantial guard intervals between the carrier locations in the frequency spectrum, which led to poor bandwidth efficiency. Allowing overlap of the adjacent spectra of individual carriers by using advanced modulation techniques (e.g., offset quadrature amplitude modulation) improved the efficiency of band usage. However the total amount of filtering required was still considerable, especially for a large number of subchannels. It was in the late 1960s when an extremely useful theoretical property was noted that yielded spectrally-efficient modem design: by letting the carriers be mutually orthogonal, the receiver can separate parallel

digital signals even if they are not bandlimited. The most attractive feature of this approach was that frequency division could be achieved not by bandpass filtering, but rather by baseband processing.

A parallel baseband processing transmission system can be realised efficiently if the discrete Fourier transform (DFT) or inverse DFT (IDFT) is employed. DFT and IDFT (which is just a complex conjugate of DFT) are implemented using computationally simple fast Fourier transform (FFT) algorithms. MC interfaces with a DFT-based architecture are known in the literature as *orthogonal frequency division multiplexing* (OFDM) and *discrete multitone* (DMT) systems. Both OFDM and DMT employ IFFT at the transmitter to form a mixture of frequency-domain data-modulated *subcarriers* (*tones*), and FFT at the receiver for separation of the data symbols. The time-domain waveforms of the subcarriers are orthogonal, yet the signal spectra corresponding to the different subcarriers overlap in frequency. Hence, the available bandwidth is used efficiently. Per subcarrier, rectangular phase shift keying (PSK) and quadrature amplitude modulation (QAM) schemes are typically used, with higher-density constellations leading to higher data rates, if symbol detection noise can be kept low.

OFDM was first commercially employed in the European Digital Audio Broadcasting (DAB) standard [3], introduced in the mid-1980s, which intended to supersede the existing analogue Amplitude and Frequency Modulation (AM, FM) systems [4]. DAB signals have approximately 1.54MHz bandwidth, and subcarriers are modulated by differential quaternary phase shift keying (DQPSK). The more recent terrestrial Digital Video Broadcasting (DVB-T) system [5], developed for the high definition digital television (HDTV) applications, is very similar to the OFDM-based DAB standard, except it uses 8MHz bandwidth, and the subcarriers can be modulated with a higher-order QAM constellation, with up to 64 points [6].

In July 1998, the IEEE 802.11 standardisation group selected OFDM as the basis for their new 802.11a Wireless Local Area Network (WLAN) standard [7]. This standard targeted a range of physical (PHY) layer data rates from 6 up to 54Mbps in the 5.2GHz band, intended for the Unlicenced National Information Infrastructure (UNII) devices [8]. This standard was the first to use OFDM in packet-based communications, in contrast to the previous broadcast-type, continuous transmission systems like DAB [3] and DVB-T [5]. In 2003, the IEEE 802.11g amendment [9] was introduced. It defines a PHY layer with analogous specifications to 802.11a (use of OFDM, modulation and coding rates from 6Mbps up to 54Mbps depending on channel conditions), but based on the 2.4GHz radio frequency (RF) carrier known as the unlicenced Industrial, Scientific and Medical (ISM) band. Parallel standardisation efforts of the ETSI Broadband Radio Access Networks (BRAN) workgroup resulted in the High-Performance LAN (HIPERLAN) Type 2 standard [10]. Both 802.11a/g and HIPERLAN/2 modems utilise constant 20MHz bandwidth and binary PSK (BPSK) to 64-QAM subcarrier modulation. These systems are optimised for the short-range indoor propagation channels, and exhibit fivefold throughput increase in comparison with the previous WLAN generation based on the single-carrier radio interfaces (e.g., IEEE 802.11b standard).

OFDM technology has also been suggested for use in *multiband ultra-wideband* (UWB) air interfaces [11], having the potential to deliver very high data rates at short ranges in arbitrary propagation environments, with only minor transmit power need and hence almost no interference with other systems. In multiband UWB radios, where ultra-short-duration information-bearing pulses are modulated by several analogue carriers with 500–800MHz wide

subbands, each analogue carrier can employ an OFDM-modulated signal. Thus, flexible and efficient utilisation of spectrum is achieved at a moderate level of complexity.

DMT defines the MC concept, which is virtually the same as OFDM, except a minor modification stipulated by the specifics of DMT application in provisioning of DSL services [12]. Unlike OFDM modems, which are designed to accommodate passband transmissions through the wireless medium, DSL transceivers have to tackle wireline baseband channels. Hence, the output of the baseband transmitter needs to be restricted to the real-valued waveform. This is achieved by mapping the conjugate modulation symbols onto the subcarriers (tones), positioned symmetrically with respect to the centre of the band, limiting the attainable bandwidth efficiency by half. Existing DMT-based high-bit-rate DSL (HDSL) and asymmetric DSL (ADSL) systems allow downlink data rates over the existing twisted-pair telephone network up to 1.6Mbps and 9Mbps respectively [13]. Very high-speed DSL (VDSL) technology, which has received considerable research attention, is regarded as the final step in the evolution of multitone data communication through legacy telephone trunks, with the promise of increasing the downlink data rate up to the theoretical maximum of 52Mbps [4][14].

The variables for a MC signal are the number of bits per symbol and the proportion of the total transmitted power that is allotted to each subcarrier. The aggregate throughput is maximised, leading to the best spectral efficiency of the MC transmissions, if these variables are chosen so that the bit error rates on all the subcarriers are equal [2]. Ideally, the optimum power distribution should be calculated by the *water-pouring* procedure based on the *channel frequency response* (CFR) weighting. *Adaptive loading* algorithms, representing a low-complexity alternative to water-pouring and implemented in DMT systems, require the receiver to measure signal-to-noise power ratios (SNR) and/or average bit error rates (BER) in the individual subbands (groups of subcarriers), to calculate the best power and bit assignments on a per-subband basis, and to send this information back to the transmitter [15]. It should be noted that in contrast to the DMT transmissions over the long-term invariant wireline channels, efficient per-subcarrier adaptive loading implementations in the mobile wireless MC modems are hampered by fast time variation of the channel, requiring significant signalling (receiver-to-transmitter feedback) overhead.

Attempts to achieve a capacity and QoS beyond what can be offered by the wired telephone channel while maintaining reasonable hardware cost have fuelled further advancements in wireless broadband system design. The use of multiple antenna technologies can enable high data rates suited for Internet and multimedia services, and increase transmission range and reliability at no loss of spectral efficiency. By employing multiple antennas, comprising the *multiple-input multiple-output* (MIMO) architecture, a set of spatial channels is created and can be used to facilitate additional data traffic subchannels, or to introduce coding redundancy in the transmissions without decreasing the throughput.

Generally, there are two broad categories of MIMO techniques. The first one aims to improve power efficiency and combat fading by utilising maximum spatial diversity that leads to extended cell coverage, longer mobile terminal battery life, and improved error-rate performance. Such techniques are known as *space-time coding*, where transmitted data streams are jointly encoded over different antennas and time slots, and include space-time block codes (STBC) and space-time trellis codes (STTC). In contrast to STTC, where decoding complexity increases

greatly with the size of modulation constellation and code length, STBC, based on orthogonal design (e.g., Alamouti code), obtains full diversity gain with low decoding complexity and therefore has been widely used. However, space-time coding is only suitable for low-data-rate services. In order to achieve very high bandwidth efficiency up to the future 10bps/Hz, the second category of MIMO systems aims to employ either *spatial multiplexing* (SM) or *transmit beamforming* (TB).

SM can significantly increase system capacity, without requiring *channel state information* (CSI) at the transmitter. It multiplexes a number of independent data streams across spatial dimensions, i.e. transmit antennas separated by location or space. At the receiver, multiple antennas and signal processing algorithms are used to separate SM streams. Separation is possible only if the number of receive antennas is greater than or equal to the number of transmit antennas (i.e., the number of independent streams). In practice, the number of receive antennas should be bigger than minimum, since this maximises likelihood of correct detection when some spatial channels appear linearly dependent, constraining the detector's degrees of freedom.

Beamforming approaches are feasible when knowledge of the channel is present at the transmitter. CSI can be acquired by the transmitter implicitly or explicitly. With explicit beamforming, the remote side sends either its channel estimates or pre-computed spatial filtering matrices to the transmit station for beamforming. With implicit beamforming, the transmitting station uses its own channel estimates as an estimate of the remote side CSI. This is based on the channel reciprocity assumption: that the channel response from the transmitting station's antenna to the remote side's antenna is the same as the channel response in the reverse path. Furthermore, depending on the acceptable explicit signalling overhead, either full CSI or partial CSI can be fed back to the transmitter. With full CSI available, the MIMO channel can be decomposed into orthogonal spatial channels commonly referred to as *eigenmodes* [16]. The larger eigenmodes are more robust to fading than the smaller ones. The transmitter can optimise power allocation and modulation over eigenmodes (rather than transmit antennas), effectively maximising throughput for a given bit error rate with virtually no spatial processing at the receiver. With partial CSI available (e.g., a data rate per antenna), receiver processing is still used to separate different spatial streams as in the SM case, but detection performance is higher than that of SM.

Recently, combinations of MC and MIMO techniques have attracted much research attention, with the emphasis on MIMO-OFDM [17][18] as a main candidate for the next generation wireless broadband interfaces, operating in the *non-line-of-sight* (NLOS), *doubly selective* propagation environments, due to its relatively low implementation complexity. The benefits of MIMO-OFDM are essential, since in the SM or TB mode the theoretical spectral efficiency is increased in proportion to the number of transmit antennas, each of which sends a unique stream of data symbols [19].

In June 2004, IEEE 802.16 BWA workgroup approved the 802.16d standard [20] defining the air interface and medium access control protocol for a wireless metropolitan area network (WMAN), intended for providing high-bandwidth wireless voice and data services for residential and enterprise use. It was the first industry-wide standard that could be used for fixed wireless access with substantially higher data rates than most existing cellular networks [21]. The 802.16 family of products, also known as WiMax by the name of the associated standardisation industry consortium, offered a low-cost (smaller in comparison with 3G solutions), quickly deployable alternative to cabled

access networks. Thus, WiMax represented a strong competitor in the high-data-rate service provisioning to locations in the world's rural and developing areas where wired broadband infrastructure was unavailable. The 802.16d devices employed a 2x1 or a 2x2 STBC-MIMO-OFDM interface, had scalable bandwidth being an integer multiple of 1.25MHz, 1.5MHz, and 1.75MHz with a maximum of 20MHz, incorporated adaptive modulation and coding modes similar to the 802.11a standard [7], and could operate in both licenced (2.5GHz and 3.5GHz) and unlicenced (5.8GHz) frequency bands.

A WiMax standard amendment IEEE 802.16e [22], published in the beginning of 2006, introduced handoff-assisted mobility support up to subscriber terminal speeds of 120km/h, mobile station power-saving management, and additional operation in the 2.3GHz band. Further refinements to the multiple-antenna technology in 802.16e (also known as mobile WiMax) specified the use of SM-MIMO in the 2x4 and 4x4 antenna configuration, as well as adaptive mode selection between SM and STBC [23]. In November 2007, mobile WiMAX was successfully adopted by ITU as one of the IMT-2000 technologies. Since then IEEE 802.16e has officially become a major global cellular wireless standard along with 3G Partnership Project's (3GPP) UMTS with High-Speed Packet Access (HSPA) and the 3GPP2 Evolution Data Optimised (EVDO) CDMA. Currently over 260 service providers are deploying fixed, nomadic and mobile WiMax networks in 110 countries [23]. Futhermore, the wireless industry has already started the evolution process toward more advanced mobile WiMax systems. This evolution includes a project at the IEEE 802.16 workgroup for developing a more advanced 802.16m standard covering almost all areas of cellular system design and maintaining backward compatibility with 802.16e.

MIMO-OFDM technology also forms the basis of the novel IEEE 802.20 standard, which is currently being developed to specify the design of cost-effective, interoperable, multi-vendor, fully mobile, high-data-rate wireless access systems, capable of working in a wide variety of licenced frequency bands and regulatory environments [24]. The 802.20 air interface is designed from the ground up, to provide seamless broadband access with up to 20MHz bandwidth for mobility classes from stationary through high-speed vehicular access in excess of 200km/h. To achieve capacity gain, both SM and TB multiantenna schemes are mentioned in the 802.20 draft. A similar proposal of utilisation of advanced antenna technologies and spatial processing is considered as part of the upcoming 3G Long-Term Evolution (LTE) [25]. This standard is targeted by 3GPP to yield throughput enhancement of the 3G network infrastructure (up to 100Mbps downlink and 50Mbps uplink over the 20MHz bandwidth), required for multimedia-oriented, packet-based data services.

Currently the IEEE 802.11n task group n is in the process of standardising the next-generation WLAN modems providing PHY data rates up to 600Mbps (according to the Draft 3.0) [26], compared to the ten-fold smaller 54Mbps in the IEEE 802.11a standard [7]. Once again, multiple antennas, namely SM, TB and STBC, are the key innovation used to obtain these benefits. In addition, there are provisions to extend the effective bandwidth to 40MHz. Note that WLAN data rates outpace the data rates available in 3G Wide Area Networks (WANs) by orders of magnitude; there are several reasons behind this. First, WLANs use wider bandwidths (dozens of MHz) that are available in unlicenced bands. Also, the low mobility and limited range requirements, as well as the need to combat smaller multipath delay spread, simplify some aspects of system design.

The benefits of increased channel capacity facilitated by MC are obvious if a single-user transmission is considered. In mobile cellular networks, the choice of *multiple access* (MA) scheme has a great impact on the achievable throughput, coverage and hardware complexity. For example, a well-known drawback of single-carrier direct sequence (DS) CDMA systems employed in 2G and 3G networks is the high computational complexity of user separation at the receiver, if the channel is time-dispersive and hence destroys orthogonality of the users in the uplink. The inherent property of parallel channeling in MC systems motivates the use of frequency division combined with time-slot division to separate the data streams of different users. The MA derivative of the OFDM scheme, *orthogonal frequency division multiple access* (OFDMA), simply allots a distinct group of subcarriers in the band to each individual user. In such a way orthogonality of the users' signals is maintained for transmission over time-dispersive channels. Thus, low computational effort is achieved at the receiver for user separation since the subcarriers are subject to independent processing. OFDMA is currently used in the WiMax standards IEEE 802.16d/e [20][22] and is proposed for the 3G LTE downlink and IEEE 802.20. A simpler MA alternative is to employ time division multiplexing (TDM) of the OFDM blocks, corresponding to different users as in the IEEE 802.11a/g/n [7][9][26] and 802.16d [20] standards.

In general, the MA schemes can be described by arbitrary precoding and spreading operations, a user-specific subcarrier allocation, and subsequent IDFT, yielding a waveform for transmission by a subscriber station. For example, conventional OFDMA uses no precoding and no spreading. The subcarrier allocation may be arbitrary, in practice typically subblock-wise (adjacent subcarriers are assigned to the same user) or interleaved (adjacent subcarriers are assigned to different users). The combination of single-user OFDM signal synthesis and spreading using Walsh-Hadamard orthogonal codes, results in MC-CDMA (pre-IDFT spreading) and MC-DS-CDMA (post-IDFT spreading) schemes [27]. Orthogonal frequency- and code-division multiplexing (OFCDM), based on the more advanced two-dimensional (2D) spreading scheme, has recently been proposed [28], but is characterised by higher latencies and instantaneous computing demand due to the per-frame receiver processing, in contrast to the per-block processing in OFDMA, MC-CDMA and MC-DS-CDMA.

In some frequency-selective channels, CFR can contain spectral nulls or deep notch regions. In the OFDMA receiver, each subcarrier is processed independently from others, hence data transmitted in the subcarrier, falling into a spectral null of the channel, cannot be recovered. This can result in continuous error bursts on such a subcarrier within the channel coherence time interval. The channel null problem can be partially solved by employing subcarrier hopping in the band, so as to distribute errors equally among the users that would ease mitigation of their effect by means of the forward error correction (FEC) techniques. However, hopping schemes impose more stringent requirements on synchronisation and coordination of user transmissions.

In MC-CDMA, MC-DS-CDMA and OFCDM, the transmitter spreads the original data stream over all subcarriers in the band. In other words, a fraction of the data symbol, corresponding to a chip of the spreading code, is transmitted through a different subcarrier. At the receiver, there is only a minimal impact of the depressed subcarriers on the recovered data symbol as its energy is scattered in the frequency domain, eliminating the aforementioned lack of frequency diversity peculiar to the uncoded OFDMA. The major challenge with the full-band spreading schemes (like MC-CDMA, MC-DS-CDMA and OFCDM) is that multiuser detection techniques are

necessary for the uplink channel because the orthogonality among users is totally distorted, resulting in MA interference (MAI). An extensive study of the MC solutions, which are resilient to MAI, while providing frequency diversity, has been carried out by Wang and Giannakis [29]. The authors proposed a new class of generalised MC (GMC) systems, employing per-subband frequency division MA, which renders independent, parallel, single-user, frequency-selective channels similarly to OFDMA, and redundant and non-redundant linear precoding, which spreads user's data symbols over the subband. Among the precoders, unitary DFT sequences have recently attracted considerable attention due to a number of useful properties. These codes are employed in the *localised* and *interleaved* FDMA (LFDMA and IFDMA) schemes, which are the preferred uplink candidates (primarily IFDMA) for 3G LTE [30]. In LFDMA, user's subcarriers are joined in subblocks, whereas in IFDMA they are scattered across the total available bandwidth, leading to higher frequency diversity. An efficient implementation of the DFT-precoded IFDMA transmitter is possible in the time domain only, without the need for IDFT processing, so that the transmitter's complexity is even lower than in OFDMA. The only disadvantage of IFDMA, in comparison with OFDMA, is reduced flexibility of adaptive transmissions. OFDMA, with full CSI at the transmitter, can be configured to schedule arbitrary subcarrier allocation and adaptive bit and power loading, maximising multiuser channel capacity and leading to high-granularity multirate services. However, adaptive OFDMA requires considerable signalling overhead, especially in the fast fading radio channels.

Finally, a coordinated space division MA (SDMA) can be supported in the uplink as, for example, defined in the IEEE 802.16e [22] and 802.20 [24] standards. In SDMA, a base station scheduler allocates several data streams on the same time-frequency slot (block/subcarrier) from several mobile transmitters [23]. Similar to the single-user SM, the base station receiver separates data streams utilising the receiver antenna array and spatial processing algorithms.

1.1.2 Properties of multiuser multiple-antenna multicarrier technology

For transmissions over wireless dispersive (multipath) media, channel-induced inter-symbol interference (ISI) is a major performance limiting factor [29]. The other is thermal noise, which is established predominantly by the low-noise amplifier in the receiver front-end.

To mitigate the dispersion effect, it is common practice to transmit data-bearing symbols in blocks. Then a *frequency-domain equalisation* approach can be adopted to compensate for the frequency-selective channel impairments at the receiver. Unlike the serial time-domain equalisers, which cannot easily handle ISI in the case of the long *channel impulse response* (CIR) and lead to noise and interference enhancement, batch frequency-domain equalisers are more robust and are characterised by lower complexity.

OFDM-like MC systems incorporate DFT at the receiver that acts as an expansion basis for the frequency-domain equalisation. To account for *inter-block interference* (IBI) induced by multipath propagation, the *cyclic prefix* (CP) is inserted at the beginning of each transmitted block. If the CP interval is greater than the maximum excess delay of the channel and time synchronisation is perfect, IBI is completely eliminated by discarding CP at the

receiver. Furthermore, CP converts the linear convolution of the transmitted signal with the multipath channel impulse response into a circular convolution (as seen by the receiver) [29]. Then the only remaining effect of the multipath is the random phase and amplitude of each subcarrier, which is easy to cope with. The received data can be recovered by tracking the frequency response (single complex-valued gain) of the subcarriers.

CP greatly simplifies frequency-domain equalisation, but at the same time increases OFDM block length, and thus reduces data rate and power efficiency. In the absence of CP, IDFT-synthesised MC signals are prone to IBI and *inter-carrier interference* (ICI). An alternative (but more complicated) approach to avoid interference is to employ filterbanks rendering subchannels with no spectral overlap (e.g., refer to [14] and [24]), though there is also an inherent bandwidth loss due to non-minimum carrier spacing.

Since OFDM(A) is the dominant type of contemporary BWA MC interfaces, all aspects of the OFDM systems and networks are experiencing considerable research attention. Herein we summarise the advantages of OFDM with respect to other transmission technologies that are of great importance from the practical standpoint:

1) High spectral efficiency, due to the minimum orthogonal separation of subcarriers.
2) Low receiver complexity, due to simple implementation by the computationally efficient FFT and splitting the frequency-selective wideband channel into the easily equalised frequency-nonselective narrowband subchannels. The non-selective character of the subchannels also yields a simple combination with MIMO.
3) Resistance to impulse noise, due to integration of the MC signal over a long block period [2].
4) High link adaptation flexibility that can maximise transmission capacity and is achieved by means of frequency-domain adaptive loading techniques (subcarrier, bit and power allocation), based on the estimated dynamic properties of the channel.
5) Low-complexity MA, based on the frequency orthogonality between users, avoiding occurrence of MAI.
6) Suitability for the broadcast services, e.g., offered in the single-frequency networks [6] where the same information is transmitted by multiple base stations, because the received signal from different synchronised sources appears as multipath propagation with long excess delays, and thus can implicitly be exploited by the OFDM receiver.
7) Enabling fractional frequency reuse (FFR) to improve cell edge performance by partitioning subcarriers available to the interfering users into several orthogonal sets [23][24].

It has been mentioned before that channels with spectral nulls or heavily faded portions of the frequency response can degrade uncoded OFDM and especially OFDMA performance. Error correction coding and interleaving techniques are known to correct single errors and error bursts, randomly occurring in different parts of the signal spectrum, at the expense of bandwidth loss due to the coding-induced data redundancy. With proper interleaving across subcarriers and consecutive blocks, and FEC where the "soft" decoding metrics are weighted by the squared channel attenuation factors, multipath turns into an OFDM system advantage by yielding frequency diversity [6]. This forms the essence of the coded OFDM (COFDM) systems employed in the existing [3][5][10][7][20] and prospective standards. Linear transmission precoding is an alternative efficient way to introduce frequency diversity and is, for instance, employed by IFDMA in the 3G LTE uplink [30]. Coding or precoding-induced frequency diversity also enables the mitigation of interfering impulses in the frequency domain

(single-tone interference). Alternatively, if the frequency location of interferers is stable and known, they can be avoided by deactivating nearby subcarriers (that also represents a form of adaptive loading) [2].

Another attractive feature of OFDM is that it allows for a smooth migration from earlier radio access technologies. In particular, fine frequency granularity, bandwidth scalability and ability to operate in vastly different spectrum allocations, makes OFDM-based 3G LTE interfaces capable of reusing spectrum resources occupied by the legacy 2G and 3G cellular networks, with no interference with other systems, exempting the service provider from applying for a new frequency band licence [25].

It can be seen that the number of advantages offered by OFDM technology are considerable. However, drawbacks exist too, though being far outnumbered by the strengths. The extent of their negative impact on system performance depends on the essence of application, with problems arising mainly in the mobile cellular networks that impose low-cost implementation, high-power efficiency and high mobility requirements on user terminals.

One serious OFDM(A) disadvantage is sensitivity to nonlinear amplification. An OFDM waveform represents a sum of many independently modulated sine-waves, and its sampled amplitude has an almost Gaussian distribution [2]. This causes large peak-to-average power ratios (PAR). Thus, the transmitter is required to have a very large linear amplification range, or distortion is experienced due to clipping of the transmitted signal when the amplifier operates close to the saturation point. The latter condition is typical for nonlinear amplifiers, which are power-efficient and cheap. On the other hand, linear power amplifiers are expensive and waste battery power in the mobile terminal, forcing the need to limit transmission range. IFDMA, which is a DFT-precoded variant of OFDMA, is characterised by a unique property among other MC schemes in its class – it has the lowest PAR, which is solely determined by the modulus of data modulation scheme and by the pulse shaping [30]. In the special case of the rectangular pulse and subcarriers modulated by the phase shift keying (PSK), IFDMA PAR is equal to unity.

Sensitivity to time and frequency synchronisation errors is the second major OFDM problem [4]. To avoid severe performance degradation, carrier frequency and timing offsets must be compensated before DFT at the receiver. Typically, frequency offsets result from Doppler shifts or oscillator imperfections, whereas timing offsets are due to the lack of synchronisation between the transmitter and the receiver. Both effects are more pronounced in the multiuser scenarios, i.e. OFDMA. Doppler shifts are inherent to any block-wise transmission through the mobile radio channel, because the block processing concept relies on the assumption that the channel is invariant during a single block and changes from block to block. This assumption is reasonable in the slow fading environments. However, rapid continuous variation of the channel response in the block period renders parasitic frequency modulation of the transmitted signal, resulting in interference between subcarriers (ICI). ICI is also the consequence of the RF carrier and timing offsets. IFDMA and OFDMA with interleaved subcarrier allocation are especially sensitive to frequency offsets in that ICI turns into stronger MAI because adjacent subcarriers belong to different users. LFDMA, which is a DFT-precoded OFDMA with the user's subcarriers bundled in the subblock, is much more robust to inter-user ICI, but exhibits over 3dB higher PAR and smaller frequency diversity than IFDMA [30]. Thus, subcarrier allocation and precoding in OFDM-based MA systems should take into account the worst impairment out of the two described.

A road towards higher spectral efficiencies in the next-generation BWA systems is paved by the potential of MIMO techniques that can be easily used in conjunction with OFDM(A). There are several critical aspects in the MIMO design, posing a general challenge for transmission capacity enhancement. First, SM-MIMO exhibits poor performance over spatially correlated channels, e.g., in the case of line-of-sight (LOS) propagation or insufficient spacing between elements in the antenna array [18]. For low-mounted base station antennas most propagation conditions are NLOS with a considerable amount of scattering, in which case the multiplexing gains of MIMO systems are very significant. It is also recommended that the distance between antenna elements is selected so as to keep antenna cross-correlation coefficients low, namely in the 0.1–0.5 range [17]. When many antennas are to be included in a compact device, space limitations can result in poor isolation between pairs of antennas and higher cross-correlation in the array that can degrade the achievable throughput of the device. Second, high integration of the RF front-end elements in the compact MIMO transceivers complicates modem implementation and increases cost [26].

Recently, MC methods have been recognised as potential candidates for the physical layer of the prospective *cognitive radio* (CR) communication systems [31]. CR will be capable of achieving maximum capacity gain of the available resources by utilising spectrum parts in the licenced bands, which are not occupied by other radio signals. At the same time CR transmissions will not interfere with the primary users of the shared spectrum. This will be achieved by adapting to the changing environmental conditions. OFDM flexibility and computational efficiency makes it a natural choice for CR [32]. The only problem with OFDM is severe out-of-band spectral leakage, which can interfere with transmissions in the adjacent bands. If this interference is regarded as unacceptable, CR air interface design should be based on the filterbank MC architectures, which employ narrow-passband filters to create non-overlapping subchannels [14][31].

1.2 Overview of multicarrier channel estimation techniques

At the (MIMO-)OFDM receiver, the baseband front-end, associated with each antenna output, implements timing synchronisation, frequency offset compensation and received power normalisation. After CP removal and DFT conversion, the signal is processed in the frequency domain. There is superposition of data symbols from different transmit antennas on each subcarrier, with each symbol having a random amplitude and phase due to attenuation and delay distortion in the channel between the corresponding transmit and receive antennas. Thus, CSI is characterised by a single complex gain per subcarrier. If CSI is available at the receiver, one can obtain the coherent estimates of transmitted symbols using these gains in the equaliser of the single-input single-output (SISO) system or detector of the MIMO system. Linear equaliser/detector coefficients can be selected based on either *zero forcing* (ZF) or *minimum mean square error* (MMSE) criterion [6]. The MMSE solution requires information about the noise power, but is more efficient as it prevents noise enhancement if CFR has spectral nulls or deep amplitude depressions.

Alternatively, it is possible to avoid coherent detection, requiring channel identification, if differential modulation is employed in the transmitter like DQPSK in the DAB standard [3]. However, differential detection incurs about 3dB performance loss in comparison with coherent detection at high SNR values, whereas for low SNR regimes and subcarriers falling into channel nulls, the penalty is well in excess of 3dB [6].

Coherent detection requires accurate CFR information for each pair of transmit and receive antennas, at each OFDM block interval. CFR acquisition can be performed by either *training-based* or *blind* channel estimation methods. In the training approach, CSI is learned from a training signal occupying a set of unmodulated subcarriers that can be transmitted in dedicated (series of) blocks (preambles) or frequency-multiplexed along with data-bearing subcarriers (e.g., refer to [5], [7], [10], [20]). In the downlink, a dedicated channel identification slot is periodically broadcast to all users. In the uplink, each user has to transmit an individual training sequence to the base station. Unlike the training-based schemes, the blind methods rely on the structural and statistical properties of the data-modulated waveform to identify channel response.

In MIMO-OFDM systems, different signals are transmitted from different antennas simultaneously, consequently the received signal is the superposition of these signals, which gives rise to channel estimation challenges. In particular, blind channel identification is not possible in the MIMO configuration without introducing redundancy into the transmit signal that may severely limit achievable capacity. Training-based channel estimators are more practical for MIMO and are similar to those employed in the SISO systems, provided that the training sequences transmitted from each antenna are orthogonal with respect to the others so that the channel from each transmit antenna can be identified uniquely.

1.2.1 Training-based channel estimation

We first consider channel identification methods relying on transmission of training information. An important subtype of training-based techniques is *pilot-assisted channel estimation*. Its principle is the acquisition of the complete CFR from only a few pilot symbols, multiplexed along with data into the transmitted sequence. These symbols can be arbitrarily positioned within the frequency band and their arrangement may change across the sequence of blocks. In order to restore the frequency response from a set of pilot observations, it is necessary to filter out the noise affecting received pilot samples, and to interpolate the channel response determined at the pilot positions onto the other active subcarriers in the band (if not all the symbols in a block are used as pilots). It should be noted that error-free interpolation can be achieved if the channel is ideally described by the lowpass model. However, permanent presence of the additive noise at the receiver input implies the use of filtering algorithms to achieve good estimation accuracy. In general, the order of *filtering* and *interpolation* operations can be arbitrary, or they can be carried out jointly. Interpolators are sometimes regarded as filters as they also perform smoothing of the input signals, filtering out highpass noise components, even if their initial task is restoration of the missing channel response parts. However, interpolation filters do not possess any optimality properties in the sense of cleaning the

estimated signal of the additive noise. Thus, filtering and interpolation stages are usually implemented separately so that they supplement each other.

The wireless channel typically exhibits high correlation in time and frequency. Filtering and interpolation algorithms, embedded into channel estimators, can capitalise on either single- (*one-dimensional*), mainly frequency-domain channel response correlation, or make full use of the channel statistics in both time and frequency domain (*two-dimensional* solutions). In the following subsubsection we will provide an overview of the existing training-based channel estimators belonging to these two classes.

1.2.1.1 One-dimensional (1D) methods

It has been shown in [33] that CFR can be described by the low-order parametric model due to the analytical relationship between the power spectral density (PSD) and the power delay profile (PDP) of the multipath channel. Hence one does not need many CFR samples to compute the unknown underlying parameters. This allows for the development of efficient frequency-domain interpolation techniques, utilising only a few pilot symbols per OFDM block.

Various types of interpolation techniques, applied to the elementary frequency-domain *least squares* (LS) estimates of CFR at the equally-spaced pilot subcarriers, have been investigated [34][35]. Here these interpolation techniques are listed in ascending performance order, namely: linear (first-order), quadratic, Gaussian, ideal lowpass (DFT-based) and cubic spline. The DFT-based interpolator has been found more prone to ICI in the continuously varying fast fading channels. It has been established by the mean square error (MSE) analysis [36] that the linear (first-order) interpolator gives a better performance gain for nearly-flat fading channels, where the noise prevails over interpolation error. However, in the case of channel environments characterised by large delay spreads, it is more preferable to adopt an interpolator based on the Fourier decomposition.

Classical interpolation techniques (e.g., polynomial, spline, etc.) are not designated specifically for noise reduction. Hence, in filtering applications they cannot be regarded as optimal solutions. For receivers employing frequency-domain signal processing, *transform-domain* estimators have been found attractive from both the complexity and performance standpoint. The principle of transform-domain techniques is to project CFR observations at the pilot subcarrier positions by means of IDFT (or alternative fast filterbank) onto (time-domain) parameter subspace of a smaller dimension, where the filtering (denoising) is performed. The final estimates are then found by the linear combinations applied to the subspace, which are implemented through the DFT (or alternative) conversion. Implementation of the DFT filterbanks is known to be very attractive from the complexity standpoint due to the availability of the FFT hardware chips in standard MC transceiver design.

Adopting the low-complexity DFT-based transform-domain architecture, Deneire et al. [37] have developed a classical *maximum likelihood* (ML) estimator, which is known as optimal (attaining the Cramer-Rao lower bound [38]) in the deterministic sense when there is no way to acquire correlation and SNR statistics of the channel. The only channel parameter, utilised by the algorithm, is the maximum excess delay, which determines the model order (parametric subspace dimension). Care should be taken when selecting its magnitude, as setting the anticipated delay too high limits the filtering gain. A reasonable value is somewhat close to the CP duration. It should be noted that in

the case of the system model affected by the additive *white Gaussian noise* (WGN), which is described by the spike correlation function of the same magnitude across the entire band, the ML solution coincides with the constrained LS (CLS) solution [39]. CLS, as the least complex and the most robust estimator, will be referred to in more detail in Chapter 3.

For the channels with a priori known CIR correlation and SNR, the transform-domain architecture can be enhanced by applying the linear MMSE filtering in the time domain [38][40]. The resultant MMSE estimator is known to be optimal in the Bayesian sense, i.e. when the prior statistical information about the unknown parameter set is available. It has been shown that the greatest performance benefit of the MMSE estimator in comparison with ML is present in the low SNR operational modes [38].

Minn and Bhargava [41] have introduced a solution with a lower complexity than the DFT-based MMSE filtering. It is based on identifying the set of the most significant (in the average power sense) CIR samples for estimation, and exhibits a minor performance loss with regard to the optimal MMSE estimator in the case of the sample-spaced multipath channels. However, the authors do not explain how to implement tracking of the selected samples. Based on a similar idea, in our related work [42], we present a modified order-recursive LS (MORLS) algorithm, which incorporates tracking functionality by picking up the most significant CIR samples while reducing the sum of squared estimation errors to the noise floor. The reason for not including this work as a part of this thesis is the minor complexity reduction with regard to the transform-domain MMSE estimator. In fact, the complexity decrease is true only for the channels with a few uncorrelated multipath components, whereas for the rich scattering environments and/or correlated impulse response samples no benefits are observed.

The frequency-domain linear MMSE (LMMSE) estimator is known to be optimal in the Bayesian sense for non-sample-spaced channels, for which the transform-domain estimators [38][40] exhibit performance deterioration due to time-domain interpolation errors in higher SNR operational modes. But implementation complexity of the frequency-domain LMMSE estimator is usually high. Edfors et al. [43] propose the use of a suboptimal estimator based on the *singular value decomposition* (SVD) of the CFR correlation matrix to save on computational effort. The works by Senol et al. [44] and Noh et al. [45] are closely related to Edfors's [43]. Assuming that the estimation subspace dimension does not exceed the CP length, there is a proposition to diagonalise the channel correlation matrix with the Karhunen-Loeve expansion, which is also known as the *eigenvalue decomposition* (EVD) [44]. Noh et al. [45] suggest lowering complexity of the frequency-domain LMMSE estimator by partitioning the channel correlation matrix into small overlapped submatrices. Then CFR estimates can be obtained, taking into account only strongly correlated subcarriers, which fall into the coherence bandwidth, and ignoring the weakly correlated ones. Similarly to the low-rank approximation methods ([43] and [44]), this scheme suffers from performance degradation in comparison with the optimal MMSE. It is also questionable whether or not this technique can be applied to the pilot-assisted systems where the estimates are constructed based on the pilot subcarriers, which can be scattered in the band with a higher spacing than the coherence bandwidth.

A common problem with [43], [44], [45] and other similar algorithms, relying on the non-Fourier decomposition, is the dependence of the expansion basis, filtering matrices and coefficients on the channel correlation function. Their low-complexity implementation is possible only when the channel correlation function is

known a priori so that it can be used in the fixed estimator design. Another problem is that the effective approximated model order should generally be selected in accordance with SNR at the receiver input so as to maximise transmission capacity. However, the study of the latter problem has not been reported so far.

In practical OFDM systems, part of the subcarriers, typically at the band edges and at DC, are left unmodulated in order to provide spectral shaping with out-of-band interference reduction, and to avoid DC offset, respectively. These unused subcarriers are termed as *virtual subcarriers* (VCs). CFR at VCs cannot be estimated directly, limiting applicability of the conventional estimators. Huang et al. [46] propose a number of modified estimators (LS and LMMSE criteria based) to address this problem.

It has already been mentioned that propagation delays of the multipath channel are generally *non-sample-spaced*, i.e. not multiple integers of a system sampling period. This results in spectral leakage (aliasing effect) after DFT processing at the receiver that lengthens the effective CIR observed in a single block and hence introduces interpolation errors if the transform-domain channel estimation is used. To resolve this problem in low-complexity DFT-based solutions, Yang et al. [47] propose a suboptimal windowed MMSE estimator design, which has only a minor performance loss in comparison with the optimal frequency-domain MMSE estimator. There is, however, a difficulty with the window shape selection as the shaping parameter is a complicated function of the noise power, and can be determined only by numerical search.

An alternative proposal is to use the discrete cosine transform (DCT) instead of DFT in the transform-domain estimator architectures [48]. CFR interpolation based on DCT allows reduction of the aliasing effect, though does not eliminate it completely. Implementation of the DCT-based estimator can be realised by fast DCT algorithms and architectures, which are competitive with FFT. Taking into account that both schemes [47] and [48] are positioned as low-complexity solutions, it could be of interest to compare their performances in the worst-case aliasing conditions.

In the work by Simeone et al. [49], the authors refer to a detailed model of the wireless channel that is parameterised as a linear combination of paths, each of which is characterised by an excess delay and complex amplitude. As direct estimation of the excess delays represents a nonlinear problem of high computational complexity, they propose tracking the delay-subspace, which can be regarded as the orthonormal basis obtained through EVD of the correlation matrix, by the recursive algorithm. The time-varying amplitudes of the multipath components are estimated by the least mean square (LMS) filter. One should point out that the computational load increases by an order of magnitude, stipulated by the QR factorisation at each recursion step of the subspace tracker, in comparison with the low-complexity systems without delay-subspace identification.

All the previously mentioned frequency and transform-domain estimation methods are directly applicable for CFR acquisition and subsequent frequency-domain equalisation in the case when the maximum excess delay of the channel does not exceed the guard interval (cyclic prefix) of the transmitted blocks. With regard to the wireless channel, this condition is typically satisfied due to the short propagation time and rapid echo attenuation of the radio signal. In the DMT and OFDM transmissions over the wireline channels, CIR length can be relatively large with respect to the MC block size that leads to a severe limitation of the bandwidth efficiency. It is known that if the prefix length is insufficient, IBI emerges between two successive blocks, as well as ICI between different subcarriers [50]. These interference effects can be eliminated by *time-domain equalisation* (TEQ), based on FIR filtering. TEQ

performs effective compression of the impulse response energy spread, based on different criteria: minimisation of the impulse response energy outside of the prefix interval [51][52], minimisation of the mean square error between the actual and desired impulse response [53], maximisation of the signal-to-interference-plus-noise ratio (SINR) at the output of equaliser [50]. Acker et al. [54] propose a per-tone equaliser (PTEQ), which is designed by transferring TEQ to the frequency domain. The main disadvantage of all TEQ approaches is significant computational complexity, especially at the equaliser's initialisation stage. A computationally efficient alternative is developed by Marelli and Fu [55]. It is based on the frequency-domain processing of the received MC signal using the fast analysis-synthesis filterbanks and subband equalisation.

A common problem of all TEQ-related methods is that they require transmission of the training sequence simultaneously on all the active subcarriers in the band. In the wireline DMT channels (e.g., xDSL trunks) characterised by the invariant state of the response, training needs to be performed only in the beginning of the communication session. In contrast to that, the necessity of the response variation tracking in the wireless channel forces frequent re-training that can be achieved by inserting pilot symbols into the transmit sequence at the selected frequencies. The inability to utilise the frequency-domain differentiation of training and data limits TEQ application in the pilot-assisted MC systems (e.g., wireless OFDM).

1.2.1.2 Two-dimensional (2D) methods

The optimal 2D MMSE estimator makes full use of the frequency (*intrablock*) and time-domain (*interblock*) correlations of CFR of the time-varying multipath fading channel. Li et al. [56] have derived the optimal 2D MMSE estimator for the OFDM system, where all subcarriers are used as a training reference. Similarly to [43], the estimator is designed through EVD and *low-rank approximation* of the frequency-domain channel correlation matrix, but the parameters comprising the subspace are subject to additional tracking in the time domain (i.e. between successive blocks). The main benefit of such an estimator architecture is the independence of the tracking filterbank's complexity on the number of subcarriers being present in the system. The authors also introduce a *robust* estimator design, which is independent of the actual channel correlation function, for the case of the sample-spaced multipath channels, and estimator implementation based on the infinite-length time-domain FIR filters. It should however be noted that the robust filters are suboptimal by their nature as they cannot achieve the true MMSE in the majority of cases.

Another related work by the same author [57] considers a 2D block-type pilot-assisted CFR interpolator, which is optimal in the MMSE sense. The author deduces the robust interpolator design, which is independent of the channel statistics. Interestingly, the robust design criteria for the interpolator are the same as for the filter considered in the previous work [56]. To reduce computational complexity of the interpolator implementation, it is proposed to use a transform-domain architecture based on 2D FFT/IFFT.

Sandell and Edfors [58] investigate application of the optimal 2D *Wiener filter* and interpolator to the pilot-assisted OFDM channel estimation. The optimal scheme is compared to the suboptimal counterparts, characterised by much smaller computational complexity that is achieved by the low-rank approximation and the use of two separable 1D filters instead of the optimal 2D architecture. The *separable filtering* approach is stipulated by the

separation property of the correlation function in the time and frequency domain [56]. Interestingly, the authors find that the estimators based on the separable filters perform even better than the optimal 2D estimator, as it appears that the latter is more susceptible to interpolation errors.

Classical interpolators have also been applied to 2D channel identification. Wang and Liu [59] propose the approximation of the doubly selective channel by means of polynomial basis functions, whereas Chang and Su [60] use polynomial regression models to express the channel response jointly in the time and frequency domain. These estimation methods do not require knowledge of the channel statistics. However, construction of the optimal polynomial estimator (with minimum model error) requires adjusting dimensions of the CFR observation window, as well as selection of the proper model order. In the work by Chang and Su [60], it is reported that performance deteriorates as the maximum Doppler frequency increases. It should also be pointed out that complexity can be higher than that of the DFT-based methods (e.g., [57]) for a large number of pilot subcarriers.

Dong et al. [61] investigate application of several kinds of 2D separable interpolators to OFDM channel estimation: lowpass-windowed sinc approach and Deslauriers-Dubuc method. Their performance is of the same order as that of the frequency-domain 2D MMSE interpolator [57], with remarkable loss in the upper SNR range due to the higher interpolation error floor, but the complexity is several orders smaller. In particular, the authors point out the drawbacks of the real-time implementation of the block-type interpolators, like [57] and [59]: large latency and memory requirements.

It should be noted that interpolators developed in the works [59], [60] and [61] are model-independent, whereas the optimal MMSE interpolator [57] relies on the underlying doubly selective channel model with a priori known time-frequency correlation function and SNR. Hence the principal advantage of the model-based MMSE interpolator (or Wiener filter) is the ability to restore 2D CFR without error in the absence of noise. This can be achieved if the number of processed samples is larger than the model order, and provided that pilot symbols are selected with the appropriate density in the time-frequency grid. On the other hand, the model-independent algorithms always produce interpolation error, but their advantage is that no prior information on channel statistics is required.

The *Kalman filter* is known as a generalised *recursive* implementation of the optimal linear MMSE estimator (Wiener filter) of infinite length, which considers time-variation of the unknown parameter(s) according to the selected dynamic model. In contrast to the Wiener filter, which is restricted to stationary signals and noise, the Kalman filter has the ability to accommodate non-stationary processes described statistically in the form of state sequences. Due to its relation to Bayesian theory, the Kalman filter has the same problems as its Wiener counterpart, i.e. it relies on the second-order statistics of signal and noise. The classical vector Kalman filter [39] can be applied to track the doubly selective channel, but it suffers from high computational complexity. Lower-complexity architecture consists of a bank of scalar Kalman filters [62]. Each of them tracks variation on a separate subcarrier, and the MMSE combiner smoothes filter outputs by taking the frequency-domain correlation into account. He and Lee [63] propose Kalman filterbank architecture, which is similar to [62]. The main difference is that the frequency-domain smoothing precedes the time variation tracking stage.

The major problems with the algorithms [62] and [63] are increased complexity in the case of a large number of subcarriers (as each of them is allocated an individual Kalman filter) and a lack of scalability when pilot-assisted

transmissions are considered. Unlike the Wiener filters, which can perform both filtering and interpolation over a fixed-length sequence of OFDM blocks, recursive implementation, inherent to the Kalman filter, processes only the current and the preceding block and therefore does not allow for interpolation of the channel response across a sequence of pilot block observations. This necessitates a separate implementation of the time-domain interpolator, bringing to naught the memory-saving benefits of the Kalman filters. It should also be noted that in contrast to the pilot interpolation, the decision-directed operation of the estimator, suggested in the works [62] and [63], is much less reliable as it is prone to uncontrolled error propagation.

Cai et al. [64] consider the same estimator structure as [56] and propose application of recursive algorithms instead of the Wiener FIR filters to track the channel response parameters in the EVD subspace. Such a tracking estimator implementation is more efficient than the approaches [62] and [63] as the complexity order is not affected by the number of subcarriers used for the reference signals. The authors point to the problem of the classical Kalman filter being sensitive to accurate information about noise variance. To overcome this challenge, they propose the use of the H_∞ filtering algorithm, which is optimal in the sense of minimising the maximum disturbance effect (instantaneous energy), instead of the Kalman estimator, which is optimal in the MMSE sense. The H_∞ and Kalman estimators have a similar structure; hence they share the same advantages and drawbacks. The main problem is still the inability of the recursive filtering structure to incorporate pilot-assisted interpolation across OFDM block sequence. The robustness extent of the H_∞ algorithm is also questionable as it is not free from the uncertainty of the design parameter selection. Furthermore, noise variance measurement is not a very big challenge, as it can be done on the virtual subcarriers [46].

All the previously mentioned channel estimation methods consider non-parameterised (or partially parameterised [49]) channel models, which are characterised by only a few deterministic and statistical properties (effective model order, correlation function and noise variance). In reality, a great variety of multipath fading channels, particular to the large cell scenarios, consist of only a few dominant propagation paths (typically two to six according to the existing ETSI and PCS channel reference models), each of which is characterised by an individual excess delay, Doppler frequency shift and complex amplitude. As the number of the underlying parameters is fixed, the channel response is uniquely identifiable through them along the observed training sequence of a certain length. This forms a basic idea behind the wireless channel sounding theory and practice [65]. Relying on such a parametric channel model, Yang et al. [66] use the minimum description length (MDL) criterion [67] to detect the number of multipath components, and employ estimation of the signal parameters by rotational invariance techniques (ESPRIT) [68] for multipath excess delay acquisition. Based on the identified set of parameters, it is further proposed to track the slow variation of the delays and the fast variation of the path gains by means of the interpath interference cancellation delay locked loops (IPIC DLLs) and the 2D Wiener filterbank (similar to [56] and [58]), respectively. Liu [69] shows how to extend ESPRIT to the 2D operation, allowing for additional identification of the Doppler frequency shift corresponding to each multipath component. The complexity of the parametric channel estimator is considerable due to the DLL bank and the EVD operation required by the ESPRIT. The results show that the performance gain, in comparison with the lower-complexity non-parametric algorithms, is achieved only when the

number of multipath components is small, otherwise the difference is minor. The latter circumstance makes this type of estimators unsuitable from the commercial standpoint for OFDM transceivers designed to operate in indoor environments with rich scattering and a multitude of propagation paths.

1.2.2 Blind channel estimation

Blind channel estimation methods avoid the use of training symbols that in theory makes them good candidates for achieving high spectral efficiency. Existing blind estimation algorithms can be classified as *statistical* and *deterministic*. The statistical approaches exploit information contained in the signal correlation matrix that is gathered over a sequence of blocks, during which the channel is assumed to remain invariant. The deterministic methods focus on some useful properties of the transmit signal, e.g., the finite size of the modulation alphabet (especially in case of BPSK and QPSK) and custom encoding at the transmitter side (e.g., differential modulation). Furthermore, some deterministic features, like linear precoding, can enhance the performance of statistical blind algorithms.

A special case of the blind estimation algorithm is referred to as *semiblind* if it relies on some minimal training information (typically a few pilot symbols), or knowledge of the channel correlation function that is insufficient for CSI recovery by conventional training-based methods.

Decision-directed (DD) estimation can be regarded as a type of semiblind approach. These techniques work in an iterative fashion: an initial estimate of the channel is computed based on the pilot symbols, or predicted from the preceding OFDM block(s). Then the estimator switches to the DD mode, in which the detected data symbols are used to improve accuracy of the channel estimate. In the algorithms proposed by Chen and Kobayashi [70] and Park et al. [71], ML criterion is used for the filtering in each iteration, employing the pilot symbols or decisions of the data symbols as a reference. Wang and Liu [72] derive a DD channel estimation algorithm, which exploits information contained in CP, in particular the property of independence of the noise samples in the prefix and the data block. The algorithm demonstrates faster convergence than its counterparts [70][71] and is able to adapt to minor variations in the channel response.

Although the iterative DD algorithms are very attractive from the complexity standpoint, they are prone to decision error propagation, leading to uncontrolled dramatic performance degradation at the end of the iterative cycle. This effect is even more likely to occur in the case of initial channel estimates being predicted from the preceding block(s), due to contribution of the prediction errors, especially in the time-varying channel conditions. Indeed, degradation of the DD receiver performance under the higher Doppler frequencies has been stressed in the work by Li [57]. When symbols for the estimator reference are taken from the decoder output (e.g., [73]), to prevent decision errors, receiver complexity and detection latency are increased dramatically as the complete decoding cycle needs to be performed at each iteration.

The problem of detection error occurrence in the DD OFDM channel tracking has been addressed by Kalyani and Giridhar [74][75]. The authors regard the erroneous decisions obtained on the deeply faded subcarriers as

outliers in the regression matrix used in the channel estimator. These outliers need to be heavily down-weighted to prevent breakdown in the fast-fading channel tracking, leading to avalanche error propagation in the case of the use of conventional channel estimation algorithms. Applying the extreme value theory (EVT) for the outlier diagnostics and the corresponding weight construction, several modified iterative estimators (LS and Kalman filter based) are proposed. Unfortunately, there is no analytical way to deduce parameters of the outlier distribution, which complicates optimal and robust estimator designs. Comprehensive analyses of the optimal EVT parameter selection, algorithm convergence (especially in the presence of ICI), impact of the modulation scheme and the number of iterations are somewhat missing in the aforementioned works, but are highly desirable from the practical standpoint.

As an alternative to the bandwidth-consuming strategy of inserting pilot symbols into the data stream, adopted in the conventional training-based systems, Ho et al. [76] and Balasubramanian et al. [77] explore the idea of *pilot embedding* when the pilot symbols are added directly to the data symbols. Thus, pilot-incurred overhead is reduced to only power allocation for the training purposes. Channel estimation algorithms, exploiting embedded pilots, work in the DD fashion. In the initial iteration, the data is treated as "noise", necessitating the estimator to filter it out using some interpolation techniques. The resultant preliminary channel estimates are used to cancel pilot-symbol interference and obtain the data estimates, which can be exploited in the subsequent iterations to achieve better accuracy. It is claimed by the authors that pilot embedding avoids catastrophic error propagation, inherent to the classical DD schemes, because embedded pilots anchor channel estimates to their true values, preventing uncontrolled deviation due to decision errors. The major concern of the proposed method is that it does not possess any optimality properties. On the one hand, it cannot deliver optimal performance achieved by means of decoupling channel estimation and detection as in the conventional pilot-assisted systems. On the other hand, there is a transmit power loss due to the added pilots. Hence, theoretically it cannot achieve the same transmission capacity as purely blind solutions not relying on any pilot information. Furthermore, the literature lacks a comprehensive comparative analysis of the DD systems with pilot insertion and pilot embedding, under the condition of optimal power allocation between training and data.

A broad family of the classical statistical blind estimation techniques are known as the *subspace-based* approach. In these algorithms the estimated correlation matrix of the received signal is processed by SVD to separate the subspace spanned by the channel-transformed signal from that spanned by only the additive noise. The inherent requirement to guarantee channel identifiability using such methods is the presence of sufficient redundancy in the transmitted signal. In the OFDM case, feasible subspace-based blind estimators have been developed relying on the use of the CP data [78] and VCs [79] to form the noise subspace. Heath and Giannakis [80] have introduced a subspace-based blind estimator exploring the cyclostationarity that CP induces to the transmitted signal. Although it is robust to the occurrence of CFR nulls on the data subcarriers, its convergence is slower than that of the classical subspace methods [78][79].

Statistical blind methods require collection of data records that are sufficiently long (typically dozens to hundreds of OFDM blocks) to render the channel output correlation matrix with adequate accuracy, assuming constant channel response. The associated estimation and detection latency might be prohibitive in practical implementations, especially in the wireless communication scenarios characterised by persistent time variation of the

channel. Deterministic blind estimation algorithms typically require fewer number of OFDM blocks. A semiblind joint channel estimation and data detection solution corresponding to the deterministic class has been proposed by Luise et al. [81] for the constant modulus signals. It requires several pilot symbols at both band edges and employs differential modulation and the Viterbi trellis decoder, which is known to be optimal in the ML sense, to detect the transmitted symbols. The detector relies on the predicted CFR using past symbol decisions (kind of DD operation) and corresponding channel estimates, with the prediction based on the low-order polynomial model, adopted for the local CFR approximation within the coherence bandwidth interval. However, implementation of the Viterbi algorithm incurs high computing complexity, exponentially increasing in the case of the higher-order modulation schemes and longer decoder memory. In a nutshell, application of the proposed joint estimator/detector is restricted to the least dispersive channels, which can be described by the low-order models, and large block sizes so as to reduce the fraction of pilot symbols. A similar idea is exploited in the work by Chang and Su [82], where the polynomial regression model [60] has been used to predict the time-varying channel. In contrast to [81], where the prediction model had preset coefficients, here the regression coefficient vector is formulated as a function of the data block of concern, thus reducing the joint estimation/detection problem to detection only. Detection is solved by minimising the quadratic LS objective function by means of linear programming, involving branch-and-bound tree search across the complex integer candidate solutions domain and recursive node metric calculation to reduce complexity. Despite a number of simplifications, the computing complexity of the detector is still very high, especially at the initialisation stage when it is recommended that pilots are used as a simpler alternative. An idea of restricting the number of search iterations at the cost of certain performance loss cannot be justified as the convergence speed of the algorithm depends on SNR and the order of the search. A similar method is considered in the work by Cui and Tellambura [83], where the joint estimation of channel and data is formulated as a mixed continuous and discrete integer LS optimisation problem, and the sphere decoding is applied to restrict the candidate solutions search space. The authors propose to improve detector's performance by taking into account the knowledge of the channel correlation and noise variance (termed as semiblind solution), making the objective function optimal in the ML sense. Similar to [82], complexity is the main problem of the algorithm and choice of the initial search radius is the biggest challenge. It should also be noted that the computing effort of the algorithms [82] and [83] grows exponentially as SNR decreases. Furthermore, even in the absence of noise, these techniques are prone to the error floor effect due to the possible mismatch of the adopted polynomial and the true channel model, leading to interpolation errors.

A principally different deterministic blind approach has been developed by Zhou and Giannakis [84], who decouple channel estimation from symbol detection rather than perform it jointly as in [81]-[83]. This method capitalises on the finite-alphabet properties of the PSK and QAM signals, namely invariance of the modulation symbol exponent's expectation irrespective of constellation point. The main weakness is computational complexity of the exhaustive integer search over the selected subcarrier set to determine the corresponding channel response phases that becomes prohibitive for high-order multipath channel models. Another performance-limiting issue is the noise enhancement when decoupling channel estimation from data detection.

Some properties of the deterministic methods can also be utilised in statistical blind processing algorithms. To alleviate faster convergence and simpler implementation of the subspace-based blind detectors, Petropulu et al. [85] propose application of a non-redundant linear precoder at the transmitter, to impose deterministic correlation structure on the signal. A generalised precoding approach, demonstrating higher processing gain at the receiver due to the more advanced analysis of the channel output correlation matrix, has been developed by Gao and Nallanathan [86]. Nevertheless, precoding-assisted blind channel estimation still requires long tracks of OFDM blocks (dozens in a sequence) to construct the signal correlation matrix.

In a nutshell, the main drawback of all blind algorithms is a considerable computational burden. The blind estimators do not exist in the closed form, opting for nonlinear processing to solve the minimax problem [79][82][83] or the homogeneous equation [80]. Hence the associated complexity is much higher than that of the most sophisticated training-based methods. Apart from that, all blind algorithms have the reference phase ambiguity problem in the case of signals drawn from symmetric constellation. This problem can be tackled by employing differential modulation, coding, or simply by inserting one-two pilot symbols into the stream, leading to a semiblind solution.

1.3 Optimisation of pilot-assisted OFDM system and related work

In parallel with research of efficient channel estimation algorithms for training-based wireless MC systems, the problem of optimal training design has recently attracted considerable attention [87]. The parasitic effects of pilot-assisted communication (reduction of spectral efficiency and useful power budget) have raised a number of questions, namely:

1) What is the optimal placement of pilot symbols in the time-frequency grid (also known as the *pilot pattern*)?

2) How many pilot symbols are needed to provide the optimal utilisation of the communication link?

3) What is the phase and power distribution amongst different pilot symbols?

4) What is the relative power allocation between pilot and data symbols?

Depending on the scope of the problem, these questions could be solved by formulating an optimisation task at different levels of the system architecture.

The early works on the pilot-assisted SISO OFDM systems [88][89] have used MSE of channel estimation as a criterion to optimise the pilot pattern. It has been found that in the fast-fading channels, pilot symbols, assigned in every block to a few subcarriers scattered in the band, lead to better performance than those assigned to all the subcarriers in the periodically transmitted blocks, assuming the same level of bandwidth efficiency. A more fundamental treatise of the MC block transmissions has been done by Ohno and Giannakis [90], who have considered the general case of the time-domain training sequence being superimposed on the data-modulated sequence. From both the complexity and performance standpoint, loading of pilot and data symbols on distinct subcarrier sets is recommended as the optimal choice for the linear type of receivers, decoupling symbol detection

from channel estimation. This is in contrast to the works [76] and [77], where the idea of adding pilot symbols to the data symbols is explored as a means to conserve bandwidth. By minimising MSE of the intrablock LS channel estimator [88]–[90], and similarly of the MMSE estimator [91], the optimal pilot pattern has been found to be *equipowered* and *equally spaced* in frequency.

A lower bound on the average channel capacity represents a more general optimisation criterion than MSE of the channel estimator or MSE of the data symbol detector. Indeed, it has been shown by Ma et al. [92] that maximising the lower bound on the channel capacity is equivalent to the minimisation of MMSE of channel estimation that is achieved only through the equispaced pilot arrangement in the case of the intrablock MMSE channel estimator [92]–[94].

According to [88], the number of pilot subcarriers has to be sufficient to recover (interpolate) CFR in the entire band without error in the absence of noise. In the case of a channel impulse response with statistically independent samples, this number corresponds to the effective CIR length. Adoption of the average capacity as a cost function for optimisation has established that the minimum sufficient number of pilots is not only optimal in terms of the bandwidth efficiency, but also in the sense of maximising the lower capacity bound [94].

When the channel is subject to slow time variation, it might appear that there is no need to transmit pilot symbols continuously, but rather it is sufficient to perform regular re-training. The optimal placement of the pilot-carrying blocks in time has been explored by Dong et al. [95], with preference given to the periodic transmission of single pilot blocks over the pilot block clusters. The work by Simeone and Spagnolini [96] can be viewed as an extension of the treatise [95] as both of them capitalise on the steady-state MSE of the Kalman filter estimating the doubly selective channel. For a given allowance on the estimation MSE, the transmitter may select the equispaced pilot pattern, which occupies the least bandwidth portion.

It should be noted that optimal training design is actually dependent on the adopted channel estimation method. The most widely spread equispaced configuration is truly optimal only for the estimators/interpolators relying on the Fourier basis. Choi and Lee [97] have shown that in the case of the general interpolators, not relying on a specific channel model, optimal positioning of pilot symbols in the time-frequency grid depends on the number of pilots, Doppler spectrum and PDP. This has also been noted by Zhang et al. [98], who discovered that in the linear interpolation assisted receiver, the clustered arrangement of the pilot subcarriers led to a better BER than the equispaced pattern.

With regard to some training parameters, MIMO-OFDM systems are equivalent to their SISO counterparts. Both Barhumi et al. [99] and Ma et al. [100] derive the MMSE-optimal pilot symbols as equipowered and equispaced in frequency. Considering the superimposed training transmissions, it is further established that pilot sequences across the different transmit antennas must be *orthogonal* [100], while Barhumi's analysis [99] leads to stricter requirement of the phase-shift orthogonal pilot sets.

Unlike the elaborated problem of the pilot placement, a few remarks should be made regarding the optimal power allocation between pilot and data symbols. First, all the related works reported above ([90]–[94] and [100]) derived the optimal power allocation, based on the assumption of an equispaced pilot pattern. However, this pattern is often impossible in practice, because of the technical limitations implied by the deactivated (virtual) subcarriers at

the band edges, as well as the need for flexible bandwidth allocations (not multiple of the number of pilots) if the OFDMA multiuser uplink transmissions are considered. Second, the closed-form expressions are presented only for the marginal cases of the high SNR, low SNR, and the independent identically distributed (IID) channel taps. As pointed out by Cai and Giannakis [91], for high SNR and IID CIR samples performance of the LS and MMSE estimators is the same, thus discouraging usage of the MMSE estimator in the analysis. In practice, the MMSE estimator can achieve a significant performance gain over the LS estimator when the channel is made up of non-sample-spaced or correlated multipath components, the number of which is typically smaller than the CIR length. This is possible only in the presence of adequate information of the second-order channel statistics, i.e. CIR correlation and SNR. This will be shown in detail in Chapter 3. Third, all the considered channel estimators have been limited to intrablock processing only. No pilot power allocation has been derived for the case of the 2D estimator, which is characterised by a noticeably smaller MSE than the intrablock (1D) MMSE estimator.

1.4 Particulars of thesis

The literature review in Section 1.2 highlights the importance of channel estimation and tracking when realising a high-data-rate wireless MC communication system, capable of approaching the maximum capacity of the non-ideal propagation medium. Summarising the main points, the emphasis in the practical system design should be on *low-complexity*, *robust*, *scalable* and *reconfigurable* estimator architectures. Comparing properties of training-based and blind channel estimation methods, it should become clear that only pilot-assisted solutions can satisfy these requirements. Indeed, the major drawback, hampering implementation of blind algorithms, is their complexity, which is several orders of magnitude higher than that of linear filtering and interpolation techniques, used in systems relying on training. Furthermore, double selectivity of the fixed and mobile radio channel necessitates frequent CSI updates that can be achieved by means of the reliable pilot-based tracking scheme, utilising temporal correlation properties of the channel rather than independent decisions of the transmitted data symbols.

It is pointed out in Section 1.3 that the performance of pilot-assisted channel estimation algorithms depends on the structure of the training sequence. Thus, channel identification cannot be regarded as a receiver-isolated problem as in the majority of works in the area. Since the objective is the overall system performance improvement, channel estimation effects should be considered jointly with detection of the power-constrained data-bearing signal, yielding a sophisticated transmission optimisation problem. With regard to pilot-assisted MIMO-OFDM optimisation, only a few results are reported in the literature, leaving this topic an open research question.

In this thesis, we propose low-complexity feasible channel estimators for SISO and MIMO-OFDM systems, suitable for a number of application scenarios. Depending on the anticipated dynamic properties of the channel, the filtering part of the estimator architecture can be easily reconfigured, allowing for a ubiquitous receiver implementation. To achieve the best symbol error rate performance, we solve the joint pilot-assisted transmitter-

receiver optimisation problem for the case of each proposed channel estimation algorithm. Contributions of the thesis are discussed in more detail in the following subsection.

1.4.1 Objectives and contributions

The starting part of the thesis is dedicated to the description of the general MIMO-OFDM system model, including doubly selective channel and basic receiver architectures. As an important part of the low-complexity receiver design, we aim to investigate several SM data symbol detection algorithms, which belong to the two classes: *linear* and *decision-feedback*. Selection of the proper detection algorithm has a direct impact on the system performance and optimisation since we adopt MSE of the detector's output as a metric to optimise power allocation between pilot and data symbols. We show that for the higher orders of receive diversity and high SNR regimes, detectors based on the ZF and MMSE criteria exhibit similar performance and hence are equivalent for system optimisation. The decision-feedback detection is closely related to the linear approach since the first (major performance-affecting) iteration output is the same as that of the linear scheme. Therefore there is no difference between these methods from the transmission optimisation standpoint. To reduce computing complexity of the receiver, we employ a suboptimal decision-feedback detector based on the *sorted QR decomposition* (SQRD). Computer simulations show that its performance is almost the same as that of the optimal V-BLAST detector, whereas the complexity is lower by an order of magnitude.

The central objective of the thesis is the design and performance analysis of efficient channel estimation algorithms. In the considered receiver model, the accuracy of channel identification is critical as the detector relies entirely on the CSI acquired by the channel estimator. Under severe channel estimation errors, recovery of the transmitted data by the detector and demodulator would simply fail. To meet the accurate CSI acquisition objective, we develop a family of the *reduced-complexity estimators*, which are suitable for both single and multiple-antenna pilot-assisted OFDM(A) systems. The estimators are based on transform-domain processing and have the same order of computational complexity, irrespective of the number of pilot subcarriers and their positioning, thus offering scalability and transceiver reconfiguration flexibility. The common estimator structure represents a cascade of successive small-dimension linear filtering modules. The number of modules, as well as their order inside the cascade, is determined by the class of the estimator (1D or 2D) and availability of the channel statistics (correlation and SNR).

We adopt an *analytical approach* towards the MSE performance assessment of the presented channel estimation algorithms. In this way we can establish the best and the worst-case estimator performance conditions and investigate filter characteristics, in particular MSE under the design mismatch.

For fine precision estimation in multipath channels, the statistics of which are not known a priori, we propose the *recursive design* of the filtering modules comprising the channel estimator. Simulation results show that in the steady state, performance of the recursive estimators approaches that of their theoretical counterparts, which are optimal in the MMSE sense.

In contrast to the majority of the channel estimators developed so far (refer to Section 1.2), our modular-type architectures are well suited for reconfigurable OFDM transceivers. Based on the observed dynamic channel properties, the appropriate filtering scheme can be chosen, and the pilot symbols can be assigned appropriate positions in the time-frequency grid.

Additionally, the thesis includes an extensive comparative analysis of the computational complexity of the developed optimal, suboptimal and recursive estimator architectures. By referring jointly to it and MSE performance results, it should become clear what estimation algorithm is more appropriate to a selected propagation environment and receiver complexity level.

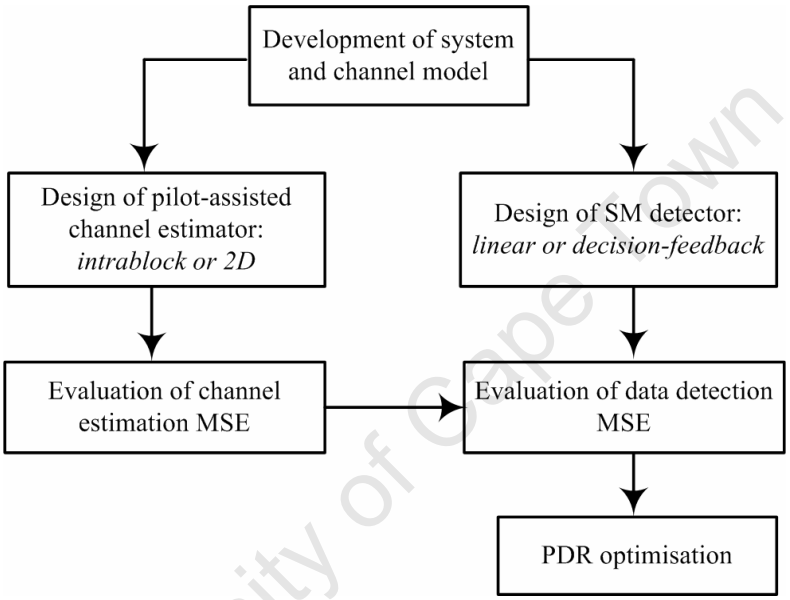


Fig.1.1. Thesis methodology

With reconfigurable channel estimation schemes, the optimal structure of pilot signal to achieve the best detection performance of the MIMO-OFDM system is an important question. The problem of optimal power allocation between the data and the pilot symbols has already been highlighted in Section 1.3. In the final part of the thesis we aim to optimise the *pilot-to-data power ratio* (PDR) for the case of developed low-complexity 1D and 2D MMSE channel estimators. Note that this section represents our main and original contribution since no attempts have been undertaken so far to optimise MC pilot structure when CSI is acquired by the 2D filtering algorithm in the multiple-antenna receiver. Furthermore, no one working in the field has utilised statistical information of the channel in optimisation. Our work fills up this knowledge gap by deriving the closed-form analytical expressions of the upper bound (suboptimal approximate value) on the optimal PDR as a function of a number of design parameters (number of subcarriers, number of pilots, number of transmit antennas, effective order of the channel model, maximum Doppler shift, SNR, etc.). The resultant PDR equations can be applied to the MIMO-OFDM systems with arbitrary (not only optimal) arrangement of the pilot subcarriers, operating in an arbitrary multipath fading channel. These properties and relatively simple functional representation of the derived analytical PDR expressions are regarded as

very useful for the reconfigurable SM-MIMO-OFDM system implementation, which is capable of adjusting transmit signal configuration according to the established channel statistics.

Fig.1.1 shows the methodology adopted in the thesis: from basic system model definition towards optimisation of PDR, subject to the selected channel estimation and detection algorithms.

1.4.2 Scope and limitations

In this thesis, we consider a conventional approach towards data symbol detection, i.e. when there is *no CSI feedback* from the receiver to the transmitter. The entire computational burden for compensating for channel distortions lies on the receiver. In the contrary case, depending on whether full or partial CSI can be fed forward to the transmitting side, adaptive loading, pre-equalisation and TB can be employed in the transmitter. If CSI is accurate, i.e. there are no signalling errors and latency is negligible, the aforementioned transmission adaptation techniques yield significant performance improvement over doubly selective channels, while allowing for very simple receiver architectures. Still practical implementations of fully adaptive systems are scarce. In mobile wireless MIMO-OFDM systems in particular, the number of adaptation parameters is large and frequent updates are necessary due to the time-varying nature of the channel. Study of the adaptive transmission methods is outside the scope of this thesis, as they represent a principally different approach to the transceiver design, and necessitate modelling of the feedback channel.

Similar to other block-wise transmission systems, the underlying design assumption is that the channel response remains constant in the interval of one block. In reality, however, the channel response varies continuously due to the Doppler effect inherent to any wireless medium. The approximation of time-varying CFR as time-invariant by means of batch processing at the receiver results in ICI that creates a detection error floor in the higher-SNR regimes. For the maximum normalised Doppler shift of 0.01, which is considered by us as a design-tolerable upper bound, the average signal-to-ICI power ratio per subcarrier is well in excess of 35dB [101]. Thus, the constant channel response assumption is reasonable, and the channel can be modelled as *block-wise fading*.

Optimised channel estimation architectures, considered in this thesis, are directly applicable to the downlink of OFDMA and linearly precoded GMC systems since the users in these MA schemes are frequency-separated. Thus, pilot and data transmissions can be regarded as channels of separate users. In the GMC systems [29][30], precoding of one user's data does not affect symbols transmitted on the other users' subcarriers. Hence, assuming that no precoding is applied to the pilot subcarriers, the functionality of the proposed channel estimation techniques is retained. Since precoded data sequences have a non-constant modulus in the frequency domain, the optimal PDR has to be reformulated with respect to the average power of the data symbol. However, in the uplink, there is a problem since each user has a different SNR and occupies only a small portion of the total bandwidth. Therefore there will be a user-specific optimal PDR due to the different number of data-bearing subcarriers and distinct channel estimation accuracy for different users.

Since we focus on optimal filtering applied to channel estimation, we do not consider the effect of CFR interpolation errors. In particular, we assume *error-free interpolation* in the frequency domain that is indeed true when CIR has a finite length and is constructed by sample-spaced multipath components, which can be correlated in general. It has been mentioned in Subsection 1.2.1 that filtering and interpolation can be regarded as independent linear problems, where interpolation errors do not depend on the additive noise, in contrast to filtering errors. Interpolation error independence on SNR implies that optimal power allocation between pilots and data should account only for the filtering gain.

Finally, we consider an *idealised spatial propagation* model, assuming that channel responses between different transmit-receive antenna pairs are statistically independent. The reader is referrred to Subsection 1.1.1 for the grounding of such a theoretical assumption, as well as a discussion of the non-ideal MIMO environments.

1.4.3 List of publications

Material presented in this thesis is partially included in the series of publications listed below in chronological order. Note that the earlier articles (I-VI) address the problem of low-complexity channel estimation, whereas the later ones (VII-X) are dedicated to the optimisation of the pilot-assisted SISO- and MIMO-OFDM systems.

I. E. Golovins and N. Ventura, "Comparative analysis of low complexity channel estimation techniques for the pilot-assisted wireless OFDM systems," in *Proceedings of the 9th South African Telecommunications and Networking Applications* ***Conference*** *(SATNAC)*, Sep. 2006.

II. E. Golovins and N. Ventura, "Low-complexity channel estimation for the wireless OFDM systems," in *Proceedings of the 12th European Wireless (EW)* ***conference***, Apr. 2007.

III. E. Golovins and N. Ventura, "Design and performance analysis of low-complexity pilot-aided OFDM channel estimators," in *Proceedings of the 6th International Workshop on Multi-Carrier and Spread Spectrum (MC-SS)*, published in ***Springer Lecture Notes*** *in Electrical Engineering*, vol. 1, May 2007.

IV. E. Golovins and N. Ventura, "Modified order-recursive least squares estimator for the noisy OFDM channels," in *Proceedings of the 5th IEEE Communications and Networking Services Research (CNSR)* ***conference***, May 2007.

V. E. Golovins and N. Ventura, "Low-complexity constrained LMMSE estimator for the sparse OFDM channels," in *Proceedings of the 8th IEEE African* ***Conference*** *(AFRICON)*, Oct. 2007.

VI. E. Golovins and N. Ventura, "Reduced-complexity recursive MMSE channel estimator for the wireless OFDM systems," in *Proceedings of the IEEE Wireless Communications and Networking* ***Conference*** *(WCNC)*, Apr. 2008.

VII. E. Golovins and N. Ventura, "Optimisation of the pilot-to-data power ratio in the MQAM-modulated OFDM systems with MMSE channel estimation," in *Proceedings of the 10th South African Telecommunications and Networking Applications* ***Conference*** *(SATNAC)*, Sep. 2007.

VIII. E. Golovins and N. Ventura, "Optimal training for the SM-MIMO-OFDM systems with MMSE channel estimation," in *Proceedings of the 6th IEEE Communications and Networking Services Research (CNSR)* ***conference***, May 2008.

IX. E. Golovins and N. Ventura, "Optimisation of the pilot-to-data power ratio in the wireless MIMO-OFDM system with low-complexity MMSE channel estimation," *Elsevier Computer Communications* ***Journal***, *Special Issue on Adaptive Multicarrier Communications and Networks*, vol. 32, pp. 465-476, Feb. 2009.

X. E. Golovins and N. Ventura, "Robust recursive two-dimensional channel estimators for the MIMO-OFDM systems," in *Proceedings of the 1st Wireless Communication Society, Vehicular Technology, Information Theory and Aerospace & Electronics Systems Technology (Wireless VITAE)* ***conference***, May 2009.

XI. E. Golovins and N. Ventura, "Performance evaluation of low-complexity decision-feedback detector for SM-MIMO-OFDM Systems," in *Proceedings of the 12th South African Telecommunications and Networking Applications* ***Conference*** *(SATNAC)*, Aug.-Sep. 2009.

XII. E. Golovins and N. Ventura, "Impact of multipath channel parameters on channel estimation performance in OFDM systems," in *Proceedings of the 9th IEEE African* ***Conference*** *(AFRICON)*, Sep. 2009.

1.5 Structure of thesis

The structure of the thesis corresponds to the adopted methodology (Fig.1.1), proceeding from basic concepts to complex system optimisation problematics. After this first introductory chapter, we define the discrete-time baseband-equivalent model of the SISO and MIMO-OFDM system, which will form the basis for further enhancements at a later stage. We dedicate substantial attention to the detailed characterisation of the doubly selective channel, emphasising its statistical properties, since understanding the channel modelling theory is important in the development of efficient channel estimation algorithms. The third section of Chapter 2 addresses SM-MIMO detector design, as the most complex receiver part in the absence of channel estimation. After discussing existing linear and non-linear detection approaches, we carry out their comparative complexity analyses and suggest a reduced-complexity suboptimal decision-feedback scheme based on SQRD as a viable alternative to the powerful, yet computationally demanding V-BLAST algorithm. Several computer simulation examples presented in Chapter 2 are designated to assess the achievable symbol error rate performance of the OFDM system in various antenna configurations when the receiver has ideal CSI.

Chapter 3 is entirely dedicated to low-complexity pilot-assisted channel estimation. We deduce linear 1D and 2D estimators satisfying a variety of properties (least complexity, robust design, recursive operation, etc.), and present their comparative complexity analysis. A useful feature of the linear estimation algorithms is the possibility of analytical evaluation of the MSE performance. We capitalise on this advantage and derive the lower and the upper MSE bounds, identifying correspondingly the best and the worst channel conditions. We also investigate the MMSE estimator's performance under the design mismatch. The chapter concludes with numerical experiments, where the

estimation MSE is calculated for several channel models, characterised by a different model order and response correlation. We also study the convergence and accuracy of the recursive algorithms when compared with their theoretical optimum counterparts.

Optimisation of the pilot-assisted transmissions is necessary since there is no trivial answer to how big bandwidth portion should be dedicated to the pilot signals, and what the relative power allocation between the pilot and the data symbols should be. We attempt to find the best suitable solutions to these problems in Chapter 4, where we investigate analytically the joint impact of different transmit and channel parameters on detection MSE. By capitalising on the upper bound on optimal PDR, we derive the suboptimal expressions of pilot-data power allocation and validate their accuracy via several numerical examples.

Finally, Chapter 5 summarises the work and provides general conclusions. We mention those parts of system implementation, which fall out of the scope of the thesis, but need to be realised in addition to the considered techniques if the main goal is transmission capacity increase.

To ease reading of the material, we include most of the extensive mathematical derivations as appendices. Appendices also contain some useful numerical examples, which are indirectly related to the material presented in the main part (e.g., illustrated analysis of the algorithm complexity). Notations and acronyms, used throughout the thesis, are defined in the corresponding pages.

2 System and Channel Model

Information transmission in any digital communication system takes place through a series of basic signal processing functions. Inclusion of specific processing blocks in the transceiver depends on the properties of the transmission channel and limitations imposed by the hardware design. In the context of the research dedicated to channel estimation and data detection, it is necessary to characterise basic distorting effects of the wireless media in order to outline a functional model, which would capture both channel-induced impairments of the received signal and an appropriate compensation mechanism recovering sent binary sequence. This model should deal only with the processes of interest, whereas idealised assumptions of other physical transmission effects can be adopted to reduce evaluation complexity and avoid misinterpretation of any results obtained through the analysis of the selected model.

This chapter introduces the discrete-time baseband-equivalent model of the OFDM system in both single-input single-output (SISO) and multiple-input multiple-output (MIMO) configuration. Section 2.1 formulates the concept of multicarrier (MC) block transmission as an efficient means to eliminate interblock interference (IBI) in case of the moderately time-varying channels, which may be regarded as time-invariant on a single block interval. The corresponding block fading channel models are discussed in Section 2.2 with explicit characterisation of the frequency and time selectivity effects. The generalised extension of the single-antenna system to the spatial multiplexing (SM) framework is described in Section 2.3, outlining basic features of the MIMO channel model, as well as detection algorithms. The chapter is concluded by a few simulation examples (Section 2.4), assessing error rate performance of the OFDM system in various configurations when the receiver has ideal channel state information (CSI).

2.1 Block-wise multicarrier transmission concept

Block-wise signal processing at the receiver greatly simplifies communication system design. The block transmission concept relies on the approximation of the fading channel as time-invariant on the interval of a single block. The most important advantage of this approximation is that a mathematical description of the channel output in a simple linear form can be achieved, which makes the application of the computationally-efficient procedures to

restore the in-block data symbols at the receiver possible. Here we present an analytical framework of a MC block transmission model. It is similar to the model introduced in the work by Ohno and Giannakis [90], but the derivation has distinctive peculiarities and we also consider variation of the channel response in-between two consecutive blocks.

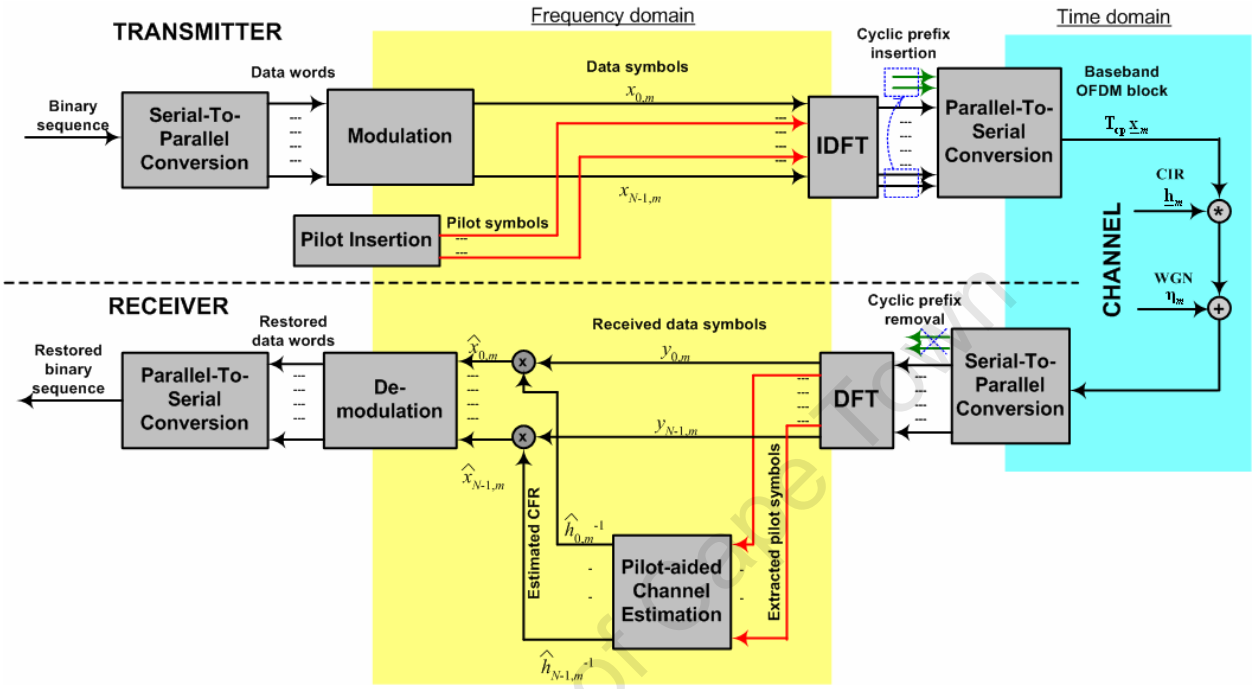


Fig.2.1. Conventional baseband OFDM system

Consider a discrete-time baseband SISO communication system that includes transmitter, receiver and an equivalent discrete-time bandlimited channel model. The transmitter and the receiver are assumed to have ideal timing and frequency synchronisation.

Let the *m*th information-carrying MC block ($m \geq 0$) be denoted as a vector, $\mathbf{x}_m = [x_{0,m} \quad \cdots \quad x_{N-1,m}]^{\mathrm{T}}$, of N complex-valued symbols, which are formed by mapping a binary sequence at the transmitter input into the points on the PSK or QAM modulation constellation (transmit signal lattice). In the MC systems this block is subject to the inverse discrete Fourier transform (IDFT), which can be written as a matrix-vector product

$$\underline{\mathbf{x}}_m = N^{-1/2}\mathbf{F}^{\mathrm{H}}\mathbf{x}_m , \qquad (2.1)$$

where $\mathbf{F}$ is the non-scaled DFT matrix [102] with the elements $[\mathbf{F}]_{k,l} = \exp(-\mathrm{j}2\pi kl/N)$, and with the property that $\mathbf{F}^{\mathrm{H}}\mathbf{F} = N\mathbf{I}$. The scaling factor $N^{-1/2}$ is introduced in (2.1) to maintain the energy of $\underline{\mathbf{x}}_m$ the same as that of $\mathbf{x}_m$.

In the physical sense, the symbols in $\mathbf{x}_m$ are loaded on the narrowband subcarriers, which are mutually orthogonal and equally spaced in the frequency domain with the minimum spacing equal to the block rate. In general, different modulation schemes can be applied simultaneously to different subcarriers, depending on what knowledge

the transmitter has about the channel. This transmission technique, called adaptive bit loading (or adaptive modulation [15]), allows for the maximisation of the channel capacity if CSI is either tracked using the bi-directional nature of the communication link or explicitly sent by the receiver to the transmitter (closed-loop signalling).

It is known that convolution of the transmit block with the impulse response of the multipath channel (also viewed as FIR filter) leads to signal dispersion in time, resulting in IBI. IBI has to be cancelled out at the receiver to enable low-complexity block-by-block processing to restore the sent data symbols. Most standards suggest to add a guard interval to the beginning of each information-carrying block to accommodate interference from the preceding block. Specifically, standardised OFDM systems [7][20] adopt the so-called *cyclic prefix* (CP) as an active guard interval, which represents a copy of the end part of $\underline{\mathbf{x}}_m$. CP insertion can be described by the product $\mathbf{T}_{\mathbf{cp}}\underline{\mathbf{x}}_m$, whereas CP removal to restore $\underline{\mathbf{x}}_m$ is performed by multiplying matrix $\mathbf{R}_{\mathbf{cp}}$ with the CP-extended vector, i.e. $\mathbf{R}_{\mathbf{cp}}\mathbf{T}_{\mathbf{cp}}\underline{\mathbf{x}}_m = \underline{\mathbf{x}}_m$, where $\mathbf{R}_{\mathbf{cp}}\mathbf{T}_{\mathbf{cp}} = \mathbf{I}_{N\times N}$.

Let the modelled discrete-time baseband-equivalent channel be characterised by an impulse response having a fixed length of L samples. This implies that the length of the guard interval must be no less than $L-1$ samples to accommodate IBI, i.e. $N_{\mathrm{cp}} \geq L-1$. Nevertheless it is recommended to set N_{cp} as close to the affordable minimum ($L-1$) as technically possible because CP samples induce redundancy, lowering the data rate, and waste transmit power.

Setting $N_{\mathrm{cp}} = L-1$ samples for the given *channel impulse response* (CIR) length L, one can express $\mathbf{R}_{\mathbf{cp}} = [\mathbf{0}_{N\times(L-1)} \quad \mathbf{I}_{N\times N}]$ and $\mathbf{T}_{\mathbf{cp}} = (\mathbf{I}_{(N+L-1)\times(N+L-1)} + \mathbf{I}_{\mathbf{cp}}^{\mathrm{T}})\mathbf{R}_{\mathbf{cp}}^{\mathrm{T}}$, where $\mathbf{I}_{\mathbf{cp}} = \left[\begin{array}{c|c} \mathbf{0}_{N\times(L-1)} & \mathbf{0}_{N\times N} \\ \hline \mathbf{I}_{(L-1)\times(L-1)} & \mathbf{0}_{(L-1)\times N} \end{array}\right]$.

Assume that the channel response is *block-wise time-invariant* (validity of this assumption has been discussed in Subsection 1.4.2), so that the value taken by CIR during transmission of $\underline{\mathbf{x}}_m$ could be denoted as $\underline{\mathbf{h}}_m = [\underline{h}_{0,m} \quad \cdots \quad \underline{h}_{L-1,m}]^{\mathrm{T}}$. On the interval of the preceding block, $\underline{\mathbf{x}}_{m-1}$, it takes a different value, $\underline{\mathbf{h}}_{m-1}$, in general. Apart from that let CIR take some intermediate value, $\underline{\mathbf{h}}_m + \boldsymbol{\Delta}_m$, during the guard interval.

The $N\times 1$ signal vector at the receiver input can be written in the following linear form:

$$\begin{aligned}\underline{\mathbf{y}}_m &= \underline{\mathbf{y}}_m^{/} + \underline{\mathbf{y}}_m^{//} + \underline{\mathbf{y}}_m^{///} + \boldsymbol{\eta}_m \\ &= \mathbf{R}_{\mathbf{cp}}(\underline{\mathbf{H}}_m^{/} + \underline{\mathbf{H}}_m^{//})\mathbf{T}_{\mathbf{cp}}\underline{\mathbf{x}}_m + \mathbf{R}_{\mathbf{cp}}\underline{\mathbf{H}}_m^{///}\mathbf{T}_{\mathbf{cp}}\underline{\mathbf{x}}_{m-1} + \mathbf{R}_{\mathbf{cp}}\boldsymbol{\eta}_m \,. \end{aligned} \tag{2.2}$$

Here $\underline{\mathbf{y}}_m^{/} = \mathbf{R}_{\mathbf{cp}}[\underline{\mathbf{h}}_m * (\mathbf{T}_{\mathbf{cp}}\underline{\mathbf{x}}_m)] = \mathbf{R}_{\mathbf{cp}}\underline{\mathbf{H}}_m^{/}\mathbf{T}_{\mathbf{cp}}\underline{\mathbf{x}}_m$ is the response of the channel on $\underline{\mathbf{x}}_m$ at its input ($*$ denotes zero-padded convolution operation, and the convolution result includes only the first $N+L-1$ elements). $\underline{\mathbf{H}}_m^{/} = \mathrm{toe}\left(\left[\underline{\mathbf{h}}_m^{\mathrm{T}} \quad \mathbf{0}_{(N-1)\times 1}^{\mathrm{T}}\right]^{\mathrm{T}}, \left[\underline{h}_{0,m} \quad \mathbf{0}_{1\times(N+L-2)}\right]\right)$, where $\mathrm{toe}(\mathbf{a},\mathbf{b}^{\mathrm{T}})$ denotes a Toeplitz matrix with the first column $\mathbf{a}$ and the first row $\mathbf{b}^{\mathrm{T}}$. The term $\underline{\mathbf{y}}_m^{//} = \mathbf{R}_{\mathbf{cp}}\underline{\mathbf{H}}_m^{//}\mathbf{T}_{\mathbf{cp}}\underline{\mathbf{x}}_m$ arises because of CIR change during the guard interval. $\underline{\mathbf{H}}_m^{//} = \left[\mathrm{toe}\left(\mathbf{I}_{(L-1)\times L}\boldsymbol{\Delta}_m, \left[\underline{\Delta}_{0,m} \quad \mathbf{0}_{1\times(N+L-2)}\right]\right)^{\mathrm{T}} \quad \mathbf{0}_{N\times(N+L-1)}^{\mathrm{T}}\right]^{\mathrm{T}}$, where $\mathbf{I}_{(L-1)\times L}$ is the rectangular identity matrix with ones

on the main diagonal and zeros elsewhere. $\underline{\mathbf{y}}_m^{///} = \mathbf{R}_{\mathbf{cp}} \underline{\mathbf{H}}_m^{///} \mathbf{T}_{\mathbf{cp}} \underline{\mathbf{x}}_{m-1}$ represents IBI. $\underline{\mathbf{H}}_m^{///} = \text{toe}\left(\mathbf{0}_{(N+L-1)\times 1}, \left[\mathbf{0}_{N\times 1}^{\mathrm{T}} \quad (\mathbf{J}_{(L-1)\times(L-1)} \mathbf{I}_{(L-1)\times L} (\underline{\mathbf{h}}_m + \boldsymbol{\Delta}_m))^{\mathrm{T}}\right]\right)$, where $\mathbf{J}_{(L-1)\times(L-1)}$ is the exchange matrix having ones on the main cross-diagonal and zeros elsewhere. $\boldsymbol{\eta}_m$ in the last term denotes additive white Gaussian noise (WGN) with $\mathrm{E}[\boldsymbol{\eta}_m] = \mathbf{0}_{(N+L-1)\times 1}$ and $\mathrm{E}[\boldsymbol{\eta}_m \boldsymbol{\eta}_m^{\mathrm{H}}] = \sigma_{\mathrm{w}}^2 \mathbf{I}_{(N+L-1)\times(N+L-1)}$.

The product $\mathbf{R}_{\mathbf{cp}} \underline{\mathbf{H}}_m^{/} \mathbf{T}_{\mathbf{cp}} = \mathbf{R}_{\mathbf{cp}} \underline{\mathbf{H}}_m^{/} (\mathbf{I} + \mathbf{I}_{\mathbf{cp}}^{\mathrm{T}}) \mathbf{R}_{\mathbf{cp}}^{\mathrm{T}}$ in (2.2) represents an $N \times N$ circulant matrix. It will be further referred to as $\underline{\mathbf{H}}_m$. This result follows from the two observations. First, $\underline{\mathbf{H}}_m^{/} \mathbf{I}_{\mathbf{cp}}^{\mathrm{T}}$ copies the first $L-1$ columns of the lower triangular matrix $\underline{\mathbf{H}}_m^{/}$ and places them over the last $L-1$ columns of the zero matrix having the same size as $\underline{\mathbf{H}}_m^{/}$ ($(N+L-1)\times(N+L-1)$). Second, the upper triangular matrix $\mathbf{R}_{\mathbf{cp}} \underline{\mathbf{H}}_m^{/} \mathbf{I}_{\mathbf{cp}}^{\mathrm{T}} \mathbf{R}_{\mathbf{cp}}^{\mathrm{T}}$, having zeros on the main diagonal and, similar to the lower triangular matrix $\mathbf{R}_{\mathbf{cp}} \underline{\mathbf{H}}_m^{/} \mathbf{R}_{\mathbf{cp}}^{\mathrm{T}} = \text{toe}\left(\left[\underline{\mathbf{h}}_m^{\mathrm{T}} \quad \mathbf{0}_{(N-L)\times 1}^{\mathrm{T}}\right]^{\mathrm{T}}, \left[\underline{h}_{0,m} \quad \mathbf{0}_{1\times(N-1)}\right]\right)$, being persymmetric in its sub-diagonal elements, supplements $\mathbf{R}_{\mathbf{cp}} \underline{\mathbf{H}}_m^{/} \mathbf{R}_{\mathbf{cp}}^{\mathrm{T}}$, yielding a matrix $\mathbf{R}_{\mathbf{cp}} \underline{\mathbf{H}}_m^{/} (\mathbf{I} + \mathbf{I}_{\mathbf{cp}}^{\mathrm{T}}) \mathbf{R}_{\mathbf{cp}}^{\mathrm{T}}$ of the right (column-wise) circulant form [103], with $\left[\underline{\mathbf{h}}_m^{\mathrm{T}} \quad \mathbf{0}_{(N-L)\times 1}^{\mathrm{T}}\right]^{\mathrm{T}}$ as the first column.

$\underline{\mathbf{H}}_m^{//}$ and $\underline{\mathbf{H}}_m^{///}$ contain non-zero elements only in the first $L-1$ rows, hence the product with $\mathbf{R}_{\mathbf{cp}}$ turns them into zero matrices and thus cancels $\underline{\mathbf{y}}_m^{//}$ and $\underline{\mathbf{y}}_m^{///}$. As a result (2.2) can be written in a simpler form of

$$\underline{\mathbf{y}}_m = \underline{\mathbf{H}}_m \underline{\mathbf{x}}_m + \mathbf{R}_{\mathbf{cp}} \boldsymbol{\eta}_m . \tag{2.3}$$

Derivation of (2.3) emphasises that one should take into account variation of the channel response only on the interval of transmission of the information block $\underline{\mathbf{x}}_m$ when selecting block length N for a given bandwidth as the system design parameter. Channel variation during the guard interval can be ignored as the corresponding part of the waveform is cancelled out. $\underline{\mathbf{x}}_m$ is often termed as the processing block, which the receiver algorithms are applied to for data recovery.

It is known [102] that all circulant matrices have the same set of eigenvectors being equal to the columns of $N^{-1/2}\mathbf{F}^{\mathrm{H}}$, and the corresponding eigenvalues are expressed as the Fourier transform of the first column of the circulant matrix. Hence the received symbols after the DFT processing can be written as

$$\begin{aligned} \mathbf{y}_m &= N^{-1/2} \mathbf{F} \underline{\mathbf{y}}_m \\ &= N^{-1/2} \mathrm{dg}(\mathbf{h}_m) \mathbf{F} \underline{\mathbf{x}}_m + N^{-1/2} \mathbf{F} \mathbf{R}_{\mathbf{cp}} \boldsymbol{\eta}_m \\ &= \mathrm{dg}(\mathbf{h}_m) \mathbf{x}_m + \mathbf{w}_m \\ &= \mathbf{X}_m \mathbf{h}_m + \mathbf{w}_m \end{aligned} , \tag{2.4}$$

where $\mathbf{h}_m = \mathbf{F} \mathbf{B} \underline{\mathbf{h}}_m$ is the non-scaled *channel frequency response* (CFR), $\mathbf{B} = \left[\mathbf{I}_{L\times L} \quad \mathbf{0}_{L\times(N-L)}\right]^{\mathrm{T}}$ is the zero-padding matrix, $\underline{\mathbf{H}}_m = N^{-1} \mathbf{F}^{\mathrm{H}} \mathrm{dg}(\mathbf{h}_m) \mathbf{F}$, $\mathrm{dg}(\mathbf{h}_m)$ and $\mathbf{X}_m = \mathrm{dg}(\mathbf{x}_m)$ are the diagonal matrices with $\mathbf{h}_m$ and $\mathbf{x}_m$ on the main

diagonal respectively, and $\mathbf{w}_m$ is the additive WGN process with the same variance as $\boldsymbol{\eta}_m$, i.e. $\mathrm{E}[\mathbf{w}_m \mathbf{w}_m^{\mathrm{H}}] = N^{-1}\mathbf{F}\mathbf{R}_{\mathbf{cp}}\mathrm{E}[\boldsymbol{\eta}_m \boldsymbol{\eta}_m^{\mathrm{H}}]\mathbf{R}_{\mathbf{cp}}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}} = \sigma_{\mathrm{w}}^2 \mathbf{I}_{N\times N}$.

As one can see, (2.4) represents a very simple linear expression of the block-by-block output for the multiplexing-type system. It has been obtained by making use of the *circular convolution* property enabled by CP insertion. This serves as an analytical grounding of an extremely important feature of the CP-assisted MC block transmissions: decoupling transmitted symbols in the frequency domain, so that they undergo channel fading independently from each other. Therefore any MC system of such type (irrespective of the modulation and coding schemes used) can be viewed as N parallel interference-free narrowband links, each of which may be subject to individual spatiotemporal processing.

When the elements of $\mathbf{x}_m$ are mutually independent random variables, and their powers and WGN variance are not known at the receiver, the optimum detector for the SISO MC block transmission described by (2.4) is based on the very fast one-tap zero-forcing (ZF) equalisation (only one multiplication per data subcarrier):

$$\hat{\mathbf{x}}_m = \mathrm{dg}(\mathbf{h}_m)^{-1}\mathbf{y}_m . \tag{2.5}$$

As the CFR $\mathbf{h}_m$ in (2.5) is typically unknown a priori, it has to be replaced by its estimate $\hat{\mathbf{h}}_m$, which can be obtained using pilot symbols. The channel estimation problem will be considered in detail in Chapter 3.

The conventional baseband SISO OFDM system with the linear receiver including pilot-assisted channel estimation and equalisation is depicted in Fig.2.1.

2.2 Channel modelling for systems with block-wise processing

Development of the optimal methods mitigating impairments caused by propagation in the wireless media is essentially preceded by the explicit characterisation of the transmission channel. Understanding of the channel-inherent processes is of crucial importance for cost-effective system design, ensuring satisfactory performance under realistic conditions. In the context of the development of an efficient channel estimator, one has to define a functional model of the propagation effects of interest and provide this model with appropriate statistical description.

In this work, the emphasis is placed on the multipath fading as the main performance-limiting phenomenon occurring in the radio links. It affects the transmitted signal in two ways: dispersion (or *frequency selectivity*) and time-variant behaviour (or *Doppler effect*), hence such environments are often referred to as "*doubly selective*". This fading class is commonly termed in the literature as the small-scale channel variation [104][105], when the received signal is subject to short-time random fluctuations caused by structural features and temporal changes of the propagation media. In this work, we do not consider shadowing (or large-scale fading), which reveals itself as space-distributed attenuation of the mean signal power, depending on the prominent terrain contour irregularities (hills, buildings, and other large-scale obstacles).

The rest of the section is organised in the following order. Subsection 2.2.1 gives a general description of the radio channel model with the structural and statistical interpretation of the underlying processes. Subsections 2.2.2 and 2.2.3 introduce and analyse components and parameters of the doubly selective channel model, which is used in the numerical performance evaluation throughout this thesis. In the last subsection we address the impact of the main channel parameters on the design of the block transmission system.

2.2.1 Wireless channel as a linear time-varying filter

It is well known that the worst scenario of radio communication is the absence of a specular path for the signal travelling between transmitter and receiver. Propagation is therefore mainly by scattering and reflection from the surfaces of terrain objects and by diffraction around them. Thus, the received multipath signal represents a superposition of different component radio waves (Fig.2.2). For densely built-up areas and indoor spaces such *non-line-of-sight* (NLOS) channel models are prevalent for determining performance limits of communication system.

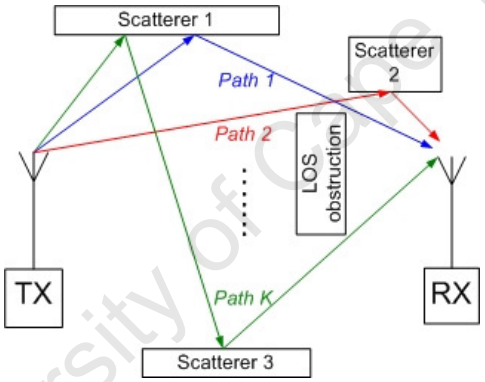


Fig.2.2. Example of NLOS multipath propagation

Apart from the received signal dispersion, its amplitude is subject to temporal variations, which are caused by the relative movement of transmitter and receiver. This movement can appear both as a transmitter/receiver terminal mobility and motion of scattering and reflecting objects in the propagation channel. It results in the time variation of the phase relationship of the incoming component waves due to the change of the individual propagation path lengths. The described phenomenon is known as the Doppler effect applied to multipath signal.

Characterisation of the majority of radio channels can be developed from general representation of the lowpass-equivalent *linear time-varying* (LTV) filter. In the analogue domain, this filter is described by the following complex-valued impulse response [104]:

$$g(\tau,t)=\sum_{k=0}^{K-1}a_k(t)\,\delta[\tau-\tau_k(t)]\,, \tag{2.6}$$

where $a_k(t)$ is the complex gain of the *k*th path, $\tau_k(t)$ is the propagation delay for the *k*th path, which generally varies in time *t*, K is the total number of propagation paths, which is equal to infinity in case of the diffuse multipath

channel (a continuum of irresolvable component waves) and has a finite value for the discrete multipath channel (relatively small number of resolvable components). Convolution with (2.6) yields response of the channel to any lowpass-equivalent signal.

The discrete multipath model is applicable mostly to rapidly changing environments like densely built-up and indoor areas [104]. It is regarded as a reference for most of the contemporary wideband terrestrial wireless systems under a reasonable approximation of constant number of multipath components K and slow variation of the delay values $\tau_k(t)$, so that they could be assumed constant too and equal to τ_k (e.g., see [106] and [107]).

Following the aforementioned arguments, for the block fading channels (2.6) is modified to

$$g(t) = \sum_{m=0}^{\infty} \sum_{k=0}^{K-1} a_{k,m}\, \delta(t - mT - \tau_k)\,, \tag{2.7}$$

where $a_{k,m}$ is the kth path gain observed in the interval of the mth block, T. The mathematical model (2.7) represents FIR filter, which can be simulated as a *tapped delay line* (TDL) with arbitrary delays and variable tap gains (Fig.2.3).

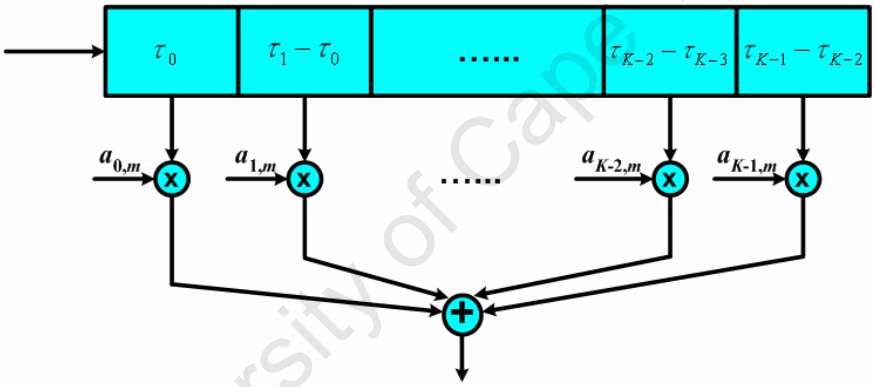


Fig.2.3. TDL model of discrete multipath fading channel

The multipath response (2.7) is defined for the continuous time. However in practice one has to deal with the discrete-time bandlimited systems. In block transmission systems with block length $N + N_{\text{cp}}$ and duration T, the effective bandwidth (inverse of the sampling rate) is equal to

$$B = (N + N_{\text{cp}})/T\,. \tag{2.8}$$

If τ_k is not an integer multiple of B^{-1}, then band-limiting yields the system-level CIR by interpolating the path gains onto the equispaced samples according to

$$\underline{\mathbf{h}}_m = \boldsymbol{\Sigma}\, \mathbf{a}_m\,, \tag{2.9}$$

where the vector $\mathbf{a}_m = [a_{k,m}]_{K\times 1}$ contains path gains, and the elements of the band-limiting matrix $\boldsymbol{\Sigma}$ are given as samples of the sinc function

$$[\boldsymbol{\Sigma}]_{l,k} = \frac{\sin \pi (l - B\tau_k)}{\pi (l - B\tau_k)}. \tag{2.10}$$

In the subsequent analysis, according to the empirical observations reported in the literature [108]-[111], we assume the primary (0th) multipath component to have the largest average power among all other components and to arrive first at the receiver with zero delay, so that the excess delays of the other paths are defined relative to the primary one. As the interpolated energy is mostly concentrated inside the main lobe of the sinc function, we can limit $l \in [0, L-1]$ in (2.10), where L is selected to ensure sufficient CIR modelling accuracy (typically encompassing more than 90% of the CIR energy). Thus, it can be said that the CIR energy in the selected structural model is strictly enclosed within the first L samples, matching the *fixed-length CIR* definition in Section 2.1.

Statistical characterisation of the random parameters τ_k and $a_{k,m}$ is presented in Subsections 2.2.2 and 2.2.3, leading to the complete definition of the channel as a LTV system. It relies on the commonly used and experimentally stipulated assumption of wideband multipath fading as the *wide sense stationary uncorrelated scattering* (WSSUS) process [65][104], which has the following interpretation. First, the impulse response of the channel filter, $g(\tau, t)$ (2.6), can be treated as a complex stochastic process, which is wide sense stationary (WSS) in time *t* and has Gaussian distribution of instantaneous values. This is satisfied for sufficiently homogenous scattering conditions [112]. Second, contributions from elemental scatterers, manifested as attenuations and phase shifts of the received signal components with different path delays, τ_k, can be assumed uncorrelated (Uncorrelated Scattering).

2.2.2 Frequency selectivity statistics

Under WSSUS assumption, the path gains $a_{k,m}$, $k \in [0, K-1]$, in (2.7) are modelled as zero-mean complex Gaussian random variables (CGRVs) produced by lowpass-filtered independent stochastic processes, which are scaled to produce the desired power, $\mathrm{E}\left[\left|a_{k,m}\right|^2\right]$, of multipath echoes. Hence the path-gain correlation matrix, $\mathbf{R}_{\mathbf{aa}} = \mathrm{E}[\mathbf{a}_m \mathbf{a}_m^{\mathrm{H}}]$, has the diagonal form. Henceforth we will term $\mathrm{dg}(\mathbf{R}_{\mathbf{aa}})$ a "*multipath intensity profile*" (MIP) to distinguish it from the channel *power-delay profile* (PDP), which is defined here in respect to the bandlimited CIR (2.9), characterised by the correlation matrix

$$\mathbf{R}_{\underline{\mathbf{hh}}} = \mathrm{E}[\underline{\mathbf{h}}_m \underline{\mathbf{h}}_m^{\mathrm{H}}] = \boldsymbol{\Sigma} \mathbf{R}_{\mathbf{aa}} \boldsymbol{\Sigma}^{\mathrm{H}}, \tag{2.11}$$

as

$$\mathbf{p} = [p_l]_{L\times 1} = \mathrm{dg}(\mathbf{R}_{\underline{\mathbf{hh}}}), \tag{2.12}$$

where, without the loss of generality, we adopt normalisation of the average channel response energy to unity, i.e. $\mathrm{tr}(\mathbf{R}_{\underline{\mathbf{hh}}}) = 1$.

For the majority of radio channels the number of multipath components, which carry most of the channel response energy, is relatively small. For example, depending on the propagation environment (indoor, outdoor,

urban, rural, hilly) the reference personal communications service (PCS) channel models [106] include 2 to 12 TDL taps, while the UHF COST 207 reference models [107] specify 4 to 12 taps.

Later multipath components go through multiple reflections more often as compared to components with shorter delays [109]. Therefore their attenuation is higher and MIP is generally characterised by non-smooth power decay [111]. The actual multipath characteristics can vary considerably, especially indoors, depending on the structure and dimensions of the building, transmitter–receiver range and presence or absence of furniture [112][113].

As stated in the literature [105], for the purpose of communication system modelling it is not necessary to take into account a variety of MIP shapes, but rather rely on the second-order channel statistics. One can use a commonly adopted characteristic – *root-mean-square* (rms) *delay spread* – in combination with simple (uniform or exponential) MIP as a measure of frequency selectivity of the channel. For such a model the *coherence bandwidth*, which defines the uniformly faded portion of frequency spectrum and hence system robustness to multipath, would be the same as for the real-world channel, for which rms delay spread was measured (e.g., see [109] and [113]-[115]). This fundamental principle of channel equivalence based on the rms delay spread greatly simplifies system modelling and allows implementing computationally optimal simulation tools. Based on PDP (2.12), the rms delay spread is formulated mathematically as the standard deviation of the excess delay:

$$\tau_{\text{rms}} = \frac{1}{B}\sqrt{\frac{\sum_{l=0}^{L-1} l^2 p_l}{\sum_{l=0}^{L-1} p_l} - \left(\frac{\sum_{l=0}^{L-1} l\, p_l}{\sum_{l=0}^{L-1} p_l}\right)^2} \quad . \tag{2.13}$$

For the outdoor environments rms delay spread increases with Tx–Rx antenna separation [109][113][115]. However indoor multipath propagation measurements [108][113] indicate that it is fairly constant and determined by the electrical properties of the building materials and transmitter and receiver surroundings, e.g., proximity of the large reflector objects.

The maximum excess delay for the indoor channels is reported to be about 200-300ns [108][116], 400ns [109] and 600ns for Tx–Rx antenna separation up to 100m [112]. The outdoor measurements [110] report that there are no excess delays beyond 1μs within a propagation range of 2km.

If the bandwidth of the communication system is sufficiently wide to allow for perfect multipath resolution (e.g., in the UWB systems), PDP and MIP coincide in all the non-zero samples. This special case is regarded as sample-spaced multipath channel, the maximum excess delay of which is equal to $B^{-1}(L-1)$. Erceg et al. [110] derive the rms delay spread for an infinite-length exponentially-decaying PDP, which models the sample-spaced channel being a good fit to the empirical experiments. In our analysis, it will be shown that for a finite-length PDP such a model suffers from inaccurate parameter definition and a more precise solution will be proposed. It should be noted that the knowledge of the PDP length is compulsory for proper receiver design to mitigate IBI efficiently. Therefore we address here the problem of the CIR/PDP length modelling for the sample-spaced channels.

2.2.2.1 *Exponentially-decaying PDP for sample-spaced multipath channels with finite-length CIR*

Define the *exponentially-decaying* PDP of the sample-spaced channel as

$$p_l = c_\alpha \mathrm{e}^{-\alpha l}, \tag{2.14}$$

where $l \in [0, L-1]$, the exponential factor $\alpha = (B\tau_0)^{-1} \geq 0$, $B^{-1}L$ represents CIR duration accommodating all the excess delays, and the coefficient $c_\alpha = \left(\sum_{l=0}^{L-1} \mathrm{e}^{-\alpha l} \right)^{-1} = \frac{1-\mathrm{e}^{-\alpha}}{1-\mathrm{e}^{-L\alpha}}$ normalises the average CIR energy to unity.

It is shown in Appendix A.1 that the normalised rms delay spread $\beta = B\tau_{\mathrm{rms}}$ and CIR length L are related according to

$$\beta = \sqrt{\frac{1}{2(\cosh\alpha - 1)} - \frac{L^2}{2(\cosh L\alpha - 1)}}. \tag{2.15}$$

The dependence (2.15) is highly nonlinear (a sum of exponents), so that provided the rms delay spread is known, the inverse function L of the argument β could be obtained only numerically by solving the nonlinear algebraic equation:

$$2\beta^2 \cosh L\alpha \cosh\alpha - (2\beta^2 + 1)\cosh L\alpha + (L^2 - 2\beta^2)\cosh\alpha + 2\beta^2 + 1 - L^2 = 0. \tag{2.16}$$

where $\alpha > 0$.

Fig.2.4 plots the function $\alpha^{-1} = B\tau_0$ of the argument $\beta = B\tau_{\mathrm{rms}}$ for a number of the PDP length values. It is seen that in contrast to the continuous infinite exponential PDP that is considered in the work [110] and is characterised by the property $\tau_0 = \tau_{\mathrm{rms}}$, in the finite-length discrete PDP, τ_0 exhibits rapid growth in the region of the relatively large rms delay spread magnitudes. It is also evident from the figure that for each given *L* there is an upper bound on the rms delay spread $\tau_{\mathrm{rms}}^{\max}$, at which τ_0 asymptotically tends to infinity (or, equivalently, $\alpha = (B\tau_0)^{-1} \to 0$). For all other $\tau_{\mathrm{rms}} < \tau_{\mathrm{rms}}^{\max}$ the condition $\tau_0 \geq \tau_{\mathrm{rms}}$ is always satisfied.

The nonlinear relationship (2.15) is considerably simplified and it becomes possible to find an inverse function of β for the two special cases, namely when $\alpha = 0$ and $L \to \infty$.

It is shown in Appendix A.1 that for $L \to \infty$, i.e. assuming the infinite-length CIR,

$$\alpha = 2\ln\left(\frac{1}{2\beta} + \sqrt{\frac{1}{4\beta^2} + 1} \right). \tag{2.17}$$

This result matches the one derived in [110].

The other scenario of interest is when τ_{rms} approaches its maximum for a PDP with the given length *L* (as shown in Fig.2.4). This results in α magnitudes in (2.14) being close to 0 and the PDP shape being almost uniform, i.e. when the average power of the CIR components is equally spread over the full range of excess delays. Thus, we determine the upper bound on τ_{rms}, which would conform to the *uniform* PDP type ($\alpha = 0$). Its derivation is presented in Appendix A.1, yielding

$$\beta_{\max} = B\,\tau_{\text{rms}}^{\max} = \frac{1}{2}\sqrt{\frac{L^2-1}{3}}\,. \tag{2.18}$$

Hence, if $\tau_{\text{rms}} = \beta B^{-1}$ is specified, then one can find the minimum length $L_{\min}$ of the CIR model that could produce such an rms delay spread magnitude. This length corresponds to the uniform PDP shape ($\alpha = 0$) and is expressed as

$$L_{\min} = \sqrt{12\beta^2 + 1}\,. \tag{2.19}$$

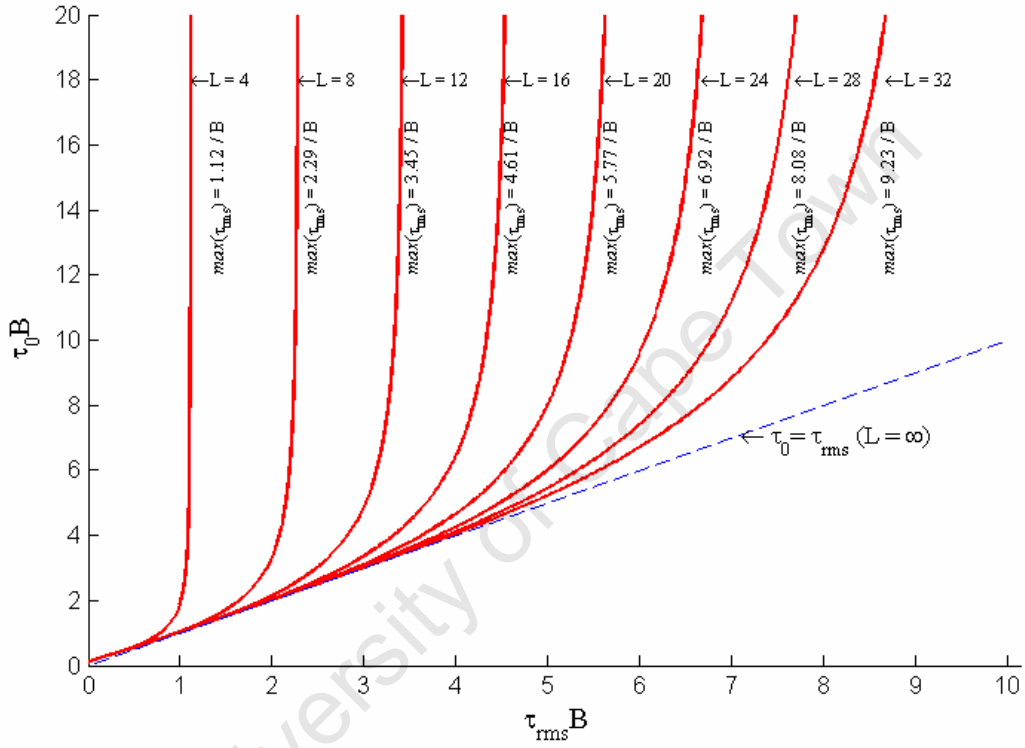


Fig.2.4. Dependence between the rms delay spread, τ_{rms}, and the exponential coefficient, τ_0, for PDPs of different lengths, L

As mentioned before, coherence bandwidth is a commonly adopted statistical criterion to characterise frequency selectivity of the channel. It describes the range of frequencies, across which received signal components exhibit strong correlation, i.e. multipath-induced gains on these frequencies are subject to only minor variation. In most cases coherence bandwidth can be considered approximately inversely proportional to the rms delay spread [114].

Frequency selectivity of the multipath channel is explicitly described by the frequency correlation function (FCF) that represents a discrete-time Fourier transform (DTFT) of PDP. For the finite-length exponentially-decaying PDP (2.14) it can be expressed (Appendix A.2) as

$$R(u) = \frac{(1-\mathrm{e}^{-\alpha})}{(1-\mathrm{e}^{-L\alpha})} \mathrm{e}^{-\frac{(L-1)(\alpha+\mathrm{j}u)}{2}} \frac{\sinh\frac{L(\alpha+\mathrm{j}u)}{2}}{\sinh\frac{(\alpha+\mathrm{j}u)}{2}}, \tag{2.20}$$

where $u \in [-\pi, \pi)$ is the angular frequency, normalised to the sampling rate.

Function (2.20) is complex in general due to the absence of evenness in (2.14). The FCF magnitude for several sample-spaced multipath channels with exponential PDP and lengths L = 16, 20 and ∞ is depicted in Fig.2.5. The coherence bandwidth is defined at the 0.9 level (other authors often define it at the 0.5 level). Note that as L increases, but τ_{rms} remains the same, coherence bandwidth tends to a slight expansion asymptotically approaching the maximum at $L = \infty$. Despite this fact, one observes approximate equality of the coherence bandwidths at the 0.9 level for a wide range of CIR lengths.

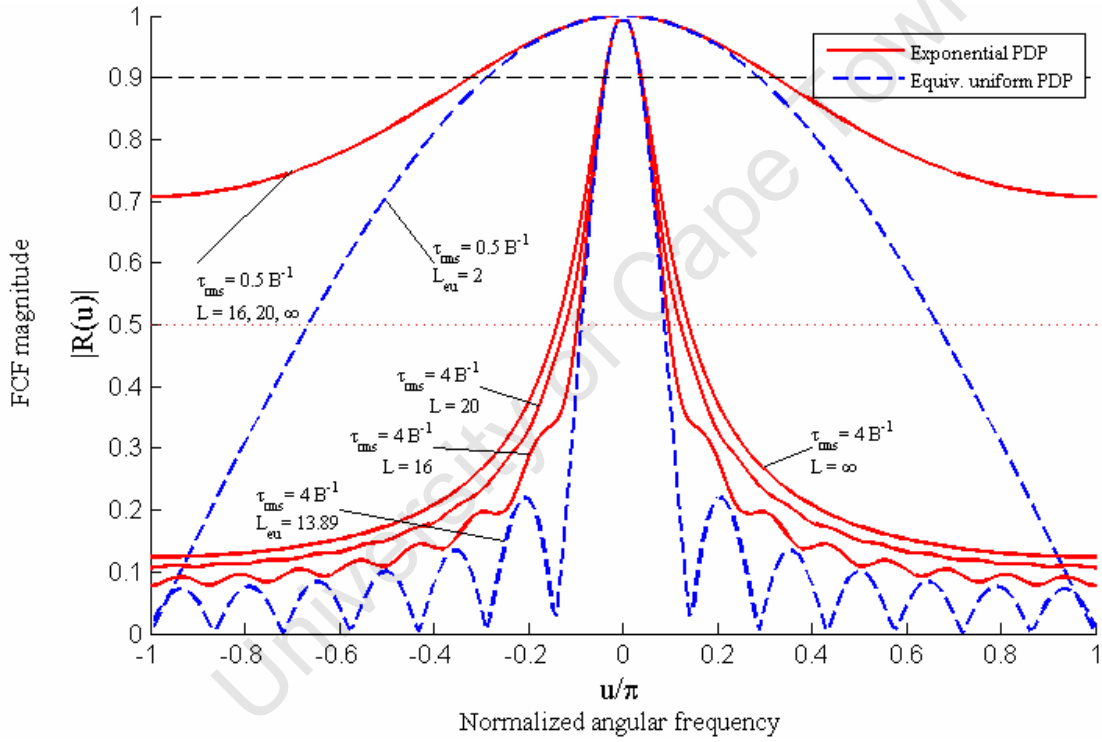


Fig.2.5. Frequency correlation function describing frequency selectivity of the sample-spaced multipath channel for several values of rms delay spread, τ_{rms}, and PDP length, L

The minimum CIR length for a specified τ_{rms} is calculated according to (2.19), and it conforms to PDP of the uniform shape ($\alpha = 0$). Uniform PDP is also termed as the case of the worst (minimum) CFR correlation [43]. Thus, it can be regarded as imposing lower bound on coherence bandwidth for a multipath channel with a given τ_{rms}. Hence we will term the minimum-length uniform PDP for a given rms delay spread as the equivalent uniform profile (also shown in Fig.2.5), for which FCF is expressed (Appendix A.2) as

$$R_u(u) = \frac{1}{L}\frac{\sin\frac{Lu}{2}}{\sin\frac{u}{2}} e^{-\frac{ju(L-1)}{2}}. \qquad (2.21)$$

From (2.21) it is obvious that shifting uniform PDP to the negative excess delay domain by $(L-1)/2$ samples produces an even function, the DTFT of which is real, and does not affect the FCF magnitude. Fig.2.6 shows the resultant FCF plotted versus the argument Lu. Although some deviation is observed on the interval [0,2π] between functions corresponding to different L values, the curves rapidly converge to the one of $L=\infty$ already for $L \geq 4$.

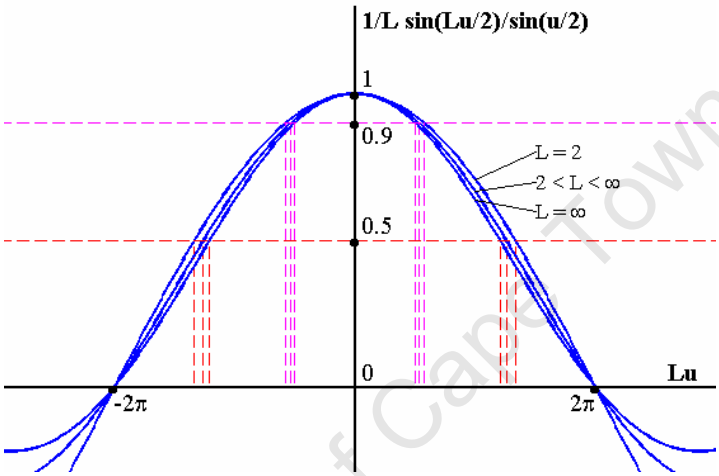


Fig.2.6. Function $\frac{1}{L}\sin\frac{Lu}{2}\Big/\sin\frac{u}{2}$ plotted versus the argument Lu

2.2.2.2 Equivalent sample-spaced two-tap MIP model

If the research objective is a performance test of the transceiver system simulation prototype without focusing on detailed channel implementation, then a simplified channel model, which would boost up speed of simulations, is desirable. The channel equivalence principle based on the rms delay spread suggests an analytically simple two-tap model. In this model, the statistical behaviour of only two components has to be simulated. We consider the sample-spaced MIP first and then describe an alternative non-sample-spaced variant.

Let rms delay spread be equal to $\tau_{\text{rms}} = \beta B^{-1}$. To reproduce an equivalent PDP satisfying τ_{rms}, the necessary condition is to have only two non-zero PDP samples, i.e.

$$p_l = \begin{cases} p_0, & l = 0 \\ p_\mu, & l = \mu \\ 0, & \text{otherwise} \end{cases}, \qquad (2.22)$$

where μB^{-1} is the excess delay of the second component relatively to the first one, and $p_0 + p_\mu = 1$ to normalise the average CIR energy to unity. We also assume decaying PDP shape, i.e. $p_0 \geq p_\mu$.

To find the unknown PDP components p_0 and p_μ we make use of (2.13):

$$\sqrt{\frac{\mu^2 p_\mu}{p_0 + p_\mu} - \left(\frac{\mu p_\mu}{p_0 + p_\mu}\right)^2} = \beta . \tag{2.23}$$

Hence solving the quadratic equation $\mu^2 p_\mu^2 - \mu^2 p_\mu + \beta^2 = 0$ with regard to p_μ yields

$$p_\mu = \frac{1}{2} - \frac{\sqrt{\mu^2 - 4\beta^2}}{2\mu}, \tag{2.24}$$

where the sample-normalised excess delay of the second component must satisfy $\mu \geq 2\beta$. The primary component is found as

$$p_0 = 1 - p_\mu = \frac{1}{2} + \frac{\sqrt{\mu^2 - 4\beta^2}}{2\mu} . \tag{2.25}$$

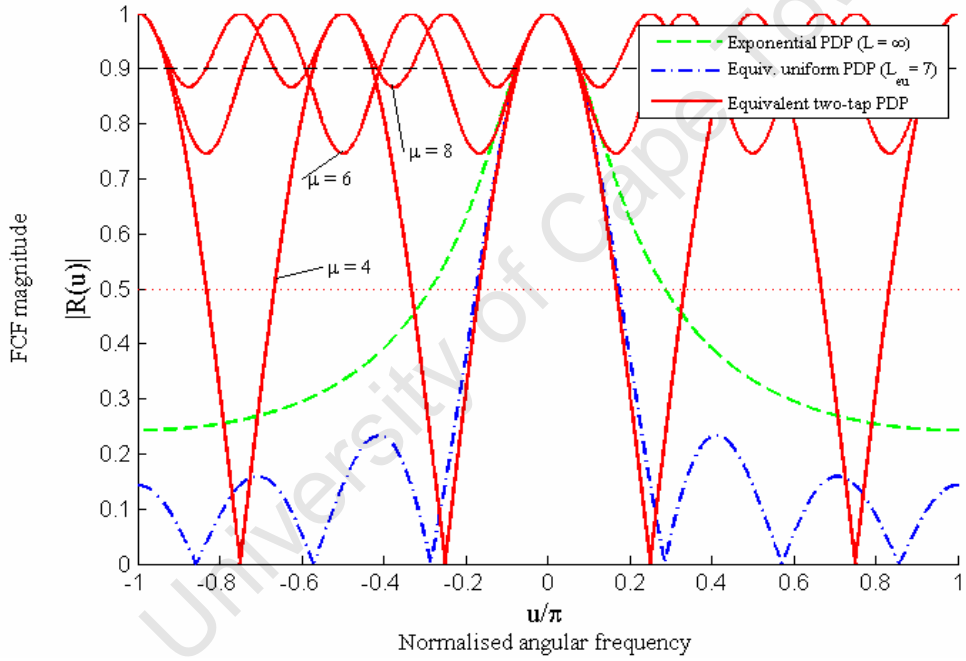


Fig.2.7. Frequency correlation function of the sample-spaced two-tap PDP model with $\tau_{\text{rms}} = 2B^{-1}$ and a range of excess delay, μB^{-1}, values

To find the coherence bandwidth of the sample-spaced two-tap channel model we derive FCF in analogy to (2.20):

$$R(u) = p_0 + p_\mu \mathrm{e}^{-\mathrm{j}u\mu}$$
$$= \sqrt{(p_0 + p_\mu \cos\mu u)^2 + p_\mu^2 \sin^2 \mu u}\ \mathrm{e}^{-\mathrm{j}\arctan\frac{p_\mu \sin\mu u}{p_0 + p_\mu \cos\mu u}}$$
$$= \sqrt{p_0^2 + 2p_0 p_\mu \cos\mu u + p_\mu^2}\ \mathrm{e}^{-\mathrm{j}\arctan\frac{\sin\mu u}{p_0/p_\mu + \cos\mu u}} . \qquad (2.26)$$

The magnitude of FCF (2.26) is plotted in Fig.2.7 for several μ values starting from the smallest, $\mu_{\min} = \lceil 2\beta \rceil$. For comparative purposes FCFs for uniform PDP and infinite-length exponential PDP, equivalent in the τ_{rms} sense, are also shown in Fig.2.7, illustrating coherence bandwidth equality at the level 0.9. The distinctive features of the two-tap model are that the sidelobes are as high as the mainlobe, and that the number of sidelobes increases with μ. The latter property makes it inappropriate for modelling frequency selectivity by an arbitrarily long CIR. Indeed, only the model with the minimum excess delay, $\mu_{\min}$, has a well-defined coherence bandwidth. The excess delay restriction might also make the two-tap PDP model unsuitable for testing channel estimators tracking the maximum excess delay of the channel that is distinguishable from the noise floor. To resolve such a problem, one should use the exponential model (2.14) instead.

2.2.2.3 Non-sample-spaced MIP models

In some cases, e.g., assessing performance of certain types of channel estimators, one needs to model the channel as non-sample-spaced. For such channels calculation of the rms delay spread according to (2.13) that uses PDP, might be quite tedious. As the band-limiting performs CIR interpolation without change of the energy spread, one may directly substitute MIP and the corresponding non-sample-spaced excess delays into (2.13).

As an example, we can formulate an equivalent non-sample-spaced two-tap MIP model similarly to (2.22), where we let equal power sharing between multipath components, $p_0 = p_\mu = 1/2$, and μ can be non-integer. Then the only excess delay satisfying (2.23) is $\mu = 2\beta$.

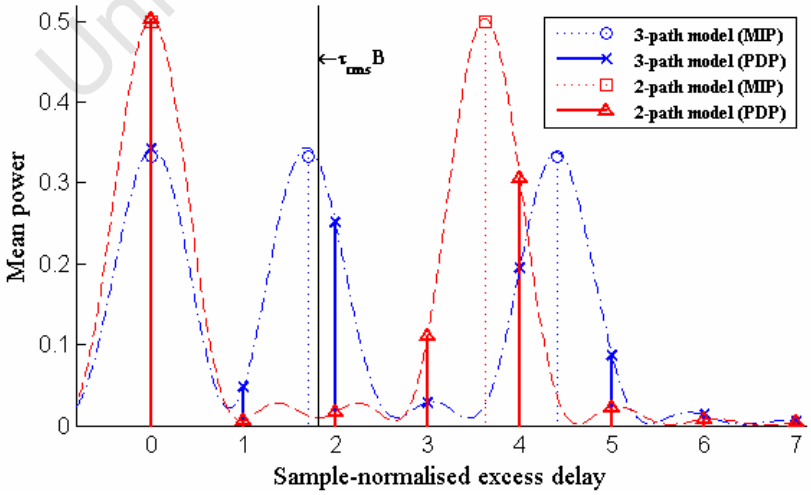


Fig.2.8. Examples of non-sample-spaced channel models

Simulation experiments described in this thesis use both *sample-spaced* and *non-sample-spaced* channel models, depending on the problem of interest. Fig.2.8 illustrates PDP examples for the non-sample-spaced channels (two and three-tap) with uniform MIPs, which are equivalent in the sense of the rms delay spread.

2.2.3 Time selectivity statistics

Whenever relative motion exists between transmitter and receiver, the lengths of the individual propagation paths vary with time. It causes the Doppler effect, which appears as a parasitic frequency modulation of the received signal. A channel, typically mobile, which is characterised by significant Doppler shifts, is called time-selective or fast fading.

It has already been mentioned in the previous subsection that the interblock temporal variation of the random path gains $a_{k,m}$ in (2.7) can be modelled by passing a WGN sequence with power $\mathrm{E}\left[\left|a_{k,m}\right|^2\right]$ for each path k, preset according to MIP, through a lowpass filter. This filter shapes the Doppler spectrum to produce the desired statistics of time variation of the channel response, which are specified by setting appropriate transfer function of the filter.

Let the modelled *autocorrelation function* (ACF) of the Doppler variation of the kth path-gain process $a_{k,m}$ be denoted as $r_{k,\nu}$, where the lag $\nu \in (-\infty, \infty)$ and $r_{k,0} = \mathrm{E}\left[\left|a_{k,m}\right|^2\right]$. The corresponding Doppler *power spectral density* (PSD),

$$S_k(\omega) = \sum_{\nu=-\infty}^{\infty} r_{k,\nu} \mathrm{e}^{-\mathrm{j}\omega\nu}\,, \tag{2.27}$$

is known as a real function of the angular frequency ω normalised to the block rate, and is strictly bandlimited in the interval $\omega \in [-\omega_{\mathrm{D}}, \omega_{\mathrm{D}})$, where $\omega_{\mathrm{D}} < \pi$ and is normally the same for $\forall k \in [0, K-1]$. Hence the transfer function of the Doppler shaping filter applied to the path-gain processes is just a square root of PSD (2.27).

The two main parameters, describing time selectivity of the channel, are the maximum angular Doppler shift, ω_{D} in (2.27), which can be defined for the vehicular communication scenario as

$$\omega_{\mathrm{D}} = 2\pi T f_{\mathrm{RF}} \mathrm{v}/\mathrm{c}\,, \tag{2.28}$$

where T is the block duration, v is the relative speed of the receiver, f_{RF} is the RF carrier frequency and c is the speed of light, and, in analogy with (2.13), the rms angular Doppler spread,

$$\omega_{\mathrm{rms}\,k} = \frac{1}{2\pi}\sqrt{\frac{\int_{-\pi}^{\pi} \omega^2 S_k(\omega)\,\mathrm{d}\omega}{\int_{-\pi}^{\pi} S_k(\omega)\,\mathrm{d}\omega} - \left(\frac{\int_{-\pi}^{\pi} \omega S_k(\omega)\,\mathrm{d}\omega}{\int_{-\pi}^{\pi} S_k(\omega)\,\mathrm{d}\omega}\right)^2}\,. \tag{2.29}$$

The latter parameter is particularly suitable for characterisation of the symmetric Doppler spectra as in such a case $\int_{-\pi}^{\pi} \omega S_k(\omega)\,\mathrm{d}\omega = 0$ in (2.29) due to the evenness of $S_k(\omega)$.

In our analysis, we consider four different Doppler models, which have been adopted in the related literature: Flat, Jakes, Gaussian and bi-Gaussian. Likewise in the majority of other works (e.g., see [58]-[60]), the Doppler spectrum has been assumed the same for all independent multipath components contributing to CIR. Hence dependence on *k* in (2.29) is excluded. However, the reader should be aware of a few UMTS reference models [107], which specify two groups of multipaths with distinct Doppler spectrums.

The Doppler PSD is found, theoretically, to be *flat* in a 3D isotropic scattering environment, where the angles of arrival are uniformly distributed in the azimuth and elevation planes [117]. It is described mathematically as

$$S_{\mathrm{F}}(\omega) = \begin{cases} \pi/\omega_{\mathrm{D}}, & \omega \in [-\omega_{\mathrm{D}}, \omega_{\mathrm{D}}) \\ 0, & \text{otherwise} \end{cases}, \tag{2.30}$$

where the total power is normalised to unity, making it equivalent to $S(\omega) = S_k(\omega)/r_{k,0}$ in (2.27). The corresponding ACF is known to be equal to

$$r_{\mathrm{F}\nu} = \sin(\omega_{\mathrm{D}}\nu)/(\omega_{\mathrm{D}}\nu), \tag{2.31}$$

and the rms Doppler spread is derived in Appendix A.3 as

$$\omega_{\mathrm{rms,F}} = \omega_{\mathrm{D}}/(2\sqrt{3}\pi). \tag{2.32}$$

The *Jakes* model, also known as dense scatterer model, is a classical Doppler spectrum shape used to describe the channel in the vehicular scenarios (receiver or transmitter moving at a certain speed) when the angles of arrival at the receiver are uniformly distributed [118]. The corresponding equations for PSD, ACF and rms Doppler spread (Appendix A.3) for the normalised Jakes model are given as follows:

$$S_{\mathrm{J}}(\omega) = \begin{cases} 2\left(\omega_{\mathrm{D}}\sqrt{1-(\omega/\omega_{\mathrm{D}})^2}\right)^{-1}, & \omega \in [-\omega_{\mathrm{D}}, \omega_{\mathrm{D}}) \\ 0, & \text{otherwise} \end{cases}, \tag{2.33}$$

$$r_{\mathrm{J}\nu} = \mathrm{J}_0(\omega_{\mathrm{D}}\nu), \tag{2.34}$$

$$\omega_{\mathrm{rms,J}} = \omega_{\mathrm{D}}/(2\sqrt{2}\pi), \tag{2.35}$$

where $\mathrm{J}_0(.)$ is Bessel function of the first kind and 0^{th} order.

It should be noted that both Flat and Jakes spectra are strictly bandlimited by the maximum Doppler shift ω_{D}. The rms spreads, (2.32) and (2.35), are proportional to ω_{D}. Hence ω_{D} provides complete parametric description for the model of the selected type. Comparing (2.32) and (2.35), it appears that the Doppler spread for the Jakes model is $\sqrt{3/2}$ times bigger than the Doppler spread of the Flat PSD. Similar to the frequency selectivity observations in Subsection 2.2.2, it is expected that transmission in the Jakes time-selective channel model would suffer from higher error rate in comparison with the Flat Doppler variation.

The *Gaussian* spectrum is considered to be a good model for multipath components with long delays in UHF communications [119]. Gaussian and Flat Doppler spectra are specified in the reference channel models for PCS applications characterised by low mobility [106]. The examples include dense urban or indoor propagation scenarios. In fixed or slowly mobile radio links, time variation of the channel response results from the relative movements of scatterers in the transmitter and receiver surroundings. It is shown in several works [120][121] that in such a case Doppler PSD is peaked at zero frequency. From the modelling standpoint the use of the Gaussian spectrum could conveniently describe this kind of time selectivity. Unlike the strictly bandlimited Flat and Jakes models, the Gaussian PSD is defined in the entire frequency range, $\omega \in [-\pi, \pi)$:

$$S_{\mathrm{G}}(\omega) = \frac{2\sqrt{\pi}}{\omega_{\mathrm{D}}} \exp\left[-\left(\frac{\omega}{\omega_{\mathrm{D}}} \right)^2 \right]. \tag{2.36}$$

Definition (2.36) has a meaningful interpretation in that it specifies PSD, which has the Doppler spread approximately equal to the Doppler spread of the Jakes PSD (2.35), i.e.

$$\omega_{\mathrm{rms,G}} \approx \omega_{\mathrm{D}} / (2\sqrt{2}\pi). \tag{2.37}$$

This is explicitly shown in Appendix A.3. The corresponding ACF is described as

$$r_{\mathrm{G}\nu} = \exp\left[-\frac{1}{4} (\omega_{\mathrm{D}} \nu)^2 \right]. \tag{2.38}$$

In contrast to the Jakes, Gaussian and Flat models, the *bi-Gaussian* spectrum is asymmetric in general. It consists of two distinct frequency-shifted Gaussian spectra and is specified in the COST 207 reference channel models to describe long echoes for urban and hilly terrain profiles [107]. The PSD and ACF expressions for the normalised bi-Gaussian model are given as

$$S_{\mathrm{B}}(\omega) = \frac{\sqrt{2\pi}}{(c_{\mathrm{B1}} + c_{\mathrm{B2}})} \left[\frac{c_{\mathrm{B1}}}{\sigma_{\mathrm{B1}}} \exp\left(-\frac{(\omega - \omega_{\mathrm{B1}})^2}{2\sigma_{\mathrm{B1}}^2} \right) + \frac{c_{\mathrm{B2}}}{\sigma_{\mathrm{B2}}} \exp\left(-\frac{(\omega - \omega_{\mathrm{B2}})^2}{2\sigma_{\mathrm{B2}}^2} \right) \right], \tag{2.39}$$

$$r_{\mathrm{B}\nu} = \frac{1}{(c_{\mathrm{B1}} + c_{\mathrm{B2}})} \left[c_{\mathrm{B1}} \exp\left(-\frac{\sigma_{\mathrm{B1}}^2 \nu^2}{2} + \mathrm{j}\omega_{\mathrm{B1}} \nu \right) + c_{\mathrm{B2}} \exp\left(-\frac{\sigma_{\mathrm{B2}}^2 \nu^2}{2} + \mathrm{j}\omega_{\mathrm{B2}} \nu \right) \right], \tag{2.40}$$

where c_{B1} and c_{B2} are unnormalised power gains of the component PSDs, ω_{B1} and ω_{B2} are the centre angular frequencies, and σ_{B1} and σ_{B2} are the corresponding standard deviations from the centre frequency. Note that ACF (2.40) is complex-valued and has odd symmetry due to PSD asymmetry.

The Doppler models described above are plotted in Fig.2.9 and Fig.2.10. We have selected "GAUS2" shape for the bi-Gaussian PSD [107], having the following parameters: $c_{\mathrm{B1}} = 1$, $c_{\mathrm{B2}} = 10^{-3/2}$, $\omega_{\mathrm{B1}} = 0.7\,\omega_{\mathrm{D}}$, $\omega_{\mathrm{B2}} = -0.4\,\omega_{\mathrm{D}}$, $\sigma_{\mathrm{B1}} = 0.1\,\omega_{\mathrm{D}}$, $\sigma_{\mathrm{B2}} = 0.15\,\omega_{\mathrm{D}}$.

One can see (Fig.2.9) that Gaussian-type spectra are not strictly bandlimited in contrast to the Flat and Jakes PSDs. Hence caution should be taken when trying to assess the maximum Doppler shift in such channels. As can be seen from Fig.2.10, the ACF magnitude span, defining the coherence time of the channel, is the widest for the bi-Gaussian spectrum. Hence this model exhibits the least time variation.

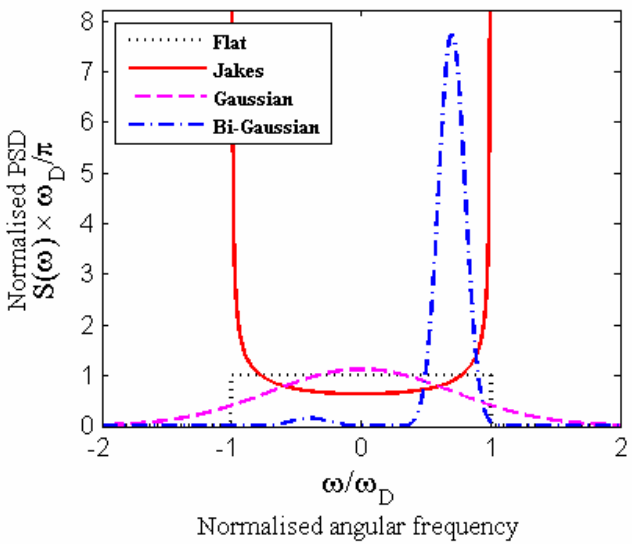


Fig.2.9. Modelled Doppler PSD

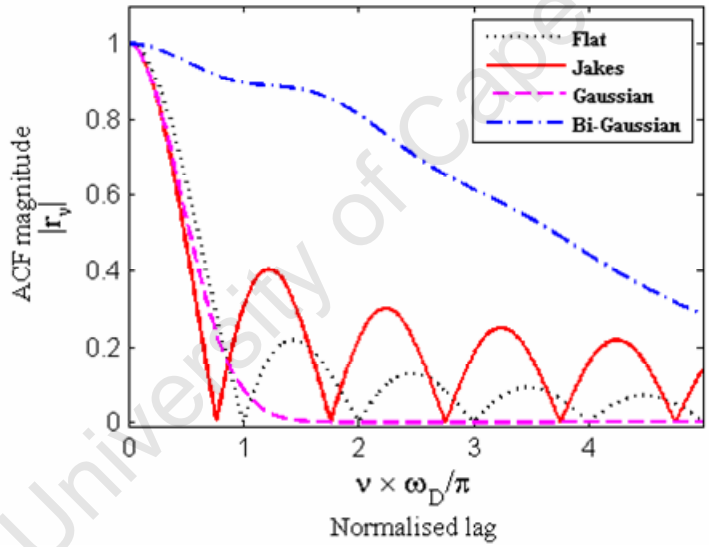


Fig.2.10. Modelled Doppler ACF magnitude (only left side plotted, right side is symmetric)

2.2.4 Impact of channel parameters on system design

Optimality of the processing algorithms incorporated in the receiver strongly depends on what channel effect is the dominant one, and how the system is configured to face this impairment (e.g., design of the channel estimator, precoder selection, etc.).

In the MC block transmission systems, the processing block length N is typically selected to be at least four times bigger than the anticipated CIR length L [20]. This rule ensures relative CFR flatness in the band of a single subcarrier and provides for realisable intrablock channel estimation techniques.

Although it is desirable to set $N >> L$ in order to have flat fading in each individual subband, the processing block duration, $B^{-1}N$, should not exceed the coherence time of the channel, i.e. the time interval, during which channel response variation is negligible, in order to avoid inter-carrier interference (ICI) and to enable low-complexity linear block-wise processing algorithms, assuming intrablock channel constancy. Reasonable system design conditions for this assumption are discussed in the works [70], [81] and [101], pointing out that the maximum Doppler shift must not exceed 0.01 of the processing block rate, $T^{-1} = B/(N + N_{\text{cp}})$, where $N_{\text{cp}} \geq L - 1$ and $N_{\text{cp}} \leq N/4$ [20]. The latter relation means that introduction of the guard interval necessitates at most 5/4 times increase of the system bandwidth to maintain the same throughput as without CP. Hence, taking into account proportionality of the Doppler bandwidth (T^{-1}) to the system bandwidth (B), 0.01 to 0.0125 of the block rate (depending on CP length) should be regarded as the maximum tolerable Doppler shift for the time-invariant CIR assumption. The *maximum angular Doppler shift* (2.28) is correspondingly restricted by

$$\omega_{\text{D}} \leq 0.025\pi . \tag{2.41}$$

The aforementioned restrictions on the processing block length can be summarised by the inequality

$$4(L-1) \leq N \leq 0.025\pi B/\omega_{\text{D}} - L + 1 . \tag{2.42}$$

2.3 Extension to MIMO configuration

Capacity of the MC block transmission system can be significantly increased if the SISO architecture described in Subsection 2.2.1 is extended to the MIMO framework. For the latter case, to enable efficient detection of the spatially multiplexed data symbols, CSI has to be identified for each *spatial layer* (SL), which is known as a unique transmit (Tx) – receive (Rx) antenna path. As the number of antennas can be quite large in general, it is therefore imperative:

1) to lessen complexity of the channel estimation algorithm;
2) to adopt a computationally fast and effective detection scheme.

In this section, extension of the discrete-time baseband OFDM system to MIMO is formulated analytically. After establishing properties of the modelled MIMO channel in Subsection 2.3.1, the received signal format, corresponding to the SM transmission mode, is defined in Subsection 2.3.2. Subsection 2.3.3 is dedicated to the overview of optimal and suboptimal detection methods used in the SM systems, accompanied by the assessment of their complexity.

2.3.1 MIMO channel features

Consider an equivalent discrete-time bandlimited channel model applied to a system with N_{tx} antennas at the transmitter and N_{rx} antennas at the receiver. For each of $N_{\text{tx}}N_{\text{rx}}$ SLs, denoted as (i, j), where $i \in [0, N_{\text{tx}} - 1]$ and $j \in [0, N_{\text{rx}} - 1]$, the channel response is described by the block-wise quasi-static approximation of the WSSUS *K*-path model (2.6):

$$g(t, j, i) = \sum_{m=0}^{\infty} \sum_{k=0}^{K-1} a_{k,m}(j,i)\, \delta[t - mT - \tau_k(j,i)], \tag{2.43}$$

where *k* and *m* are correspondingly the path and the block indices, $\tau_k(j,i)$ is the path delay, $T = B^{-1}(N + N_{\text{cp}})$ is the block duration, path gains $\mathbf{a}_m(j,i) = [a_{k,m}(j,i)]_{K\times 1}$ represent zero-mean CGRVs produced by the MIP and Doppler-shaped independent stochastic processes as described in Section 2.2.

The bandlimited CIR, $\underline{\mathbf{h}}_m(j,i) = [\underline{h}_{l,m}(j,i)]_{L\times 1}$, $l \in [0, L-1]$, corresponding to the (i, j)th SL, is linked to (2.43) according to the formula

$$\underline{\mathbf{h}}_m(j,i) = \mathbf{\Sigma}(j,i)\, \mathbf{a}_m(j,i), \tag{2.44}$$

where, in analogy to the argument presented in Subsection 2.2.1, we assume a strict CIR energy enclosure within the first L samples ($L \le N_{\text{cp}} + 1$), and the elements of the band-limiting matrix $\mathbf{\Sigma}(j,i)$ are given as

$$[\mathbf{\Sigma}(j,i)]_{l,k} = \frac{\sin \pi[l - B\tau_k(j,i)]}{\pi[l - B\tau_k(j,i)]}. \tag{2.45}$$

By the proper antenna array design, in particular setting spacing between the antenna elements to be not less than a quarter-wavelength [65], it can be assumed that channels at different SLs are *homogeneous* [16], implying non-correlatedness of their responses, i.e. $\mathrm{E}[\underline{\mathbf{h}}_m(j_1,i_1)\underline{\mathbf{h}}_m(j_2,i_2)^{\mathrm{H}}] = \mathbf{0}$ for $\forall i_1, i_2 \in [0, N_{\text{tx}} - 1]$, $\forall j_1, j_2 \in [0, N_{\text{rx}} - 1]$, where $i_1 \neq i_2$ or/and $j_1 \neq j_2$. Furthermore, we let the CIR correlation matrix $\mathbf{R}_{\underline{\mathbf{hh}}}(j,i) = \mathrm{E}[\underline{\mathbf{h}}_m(j,i)\underline{\mathbf{h}}_m(j,i)^{\mathrm{H}}]$ be identical for all SLs (denoted simply as $\mathbf{R}_{\underline{\mathbf{hh}}}$). The latter assumption holds reasonably well if antenna spacing inside the array is small in comparison with propagation distance, so that the multipath environment can be considered specular, with fixed scattering delays [16], i.e. τ_k and $\mathbf{\Sigma}$ in (2.43)-(2.45) are SL-independent. The average energy of the channel response is hereafter normalised to unity, i.e. $\mathrm{tr}(\mathbf{R}_{\underline{\mathbf{hh}}}) = 1$.

2.3.2 SM-MIMO-OFDM system model

We consider a discrete-time baseband spatially multiplexed MIMO-OFDM system (Fig.2.11) with the number of Rx antennas being no less than the number of Tx antennas, i.e. $N_{\text{rx}} \ge N_{\text{tx}}$. This condition is stipulated by the restrictions of the linear SM model, which must guarantee error-free symbol detection in the absence of noise at the receiver input (see next subsection for details).

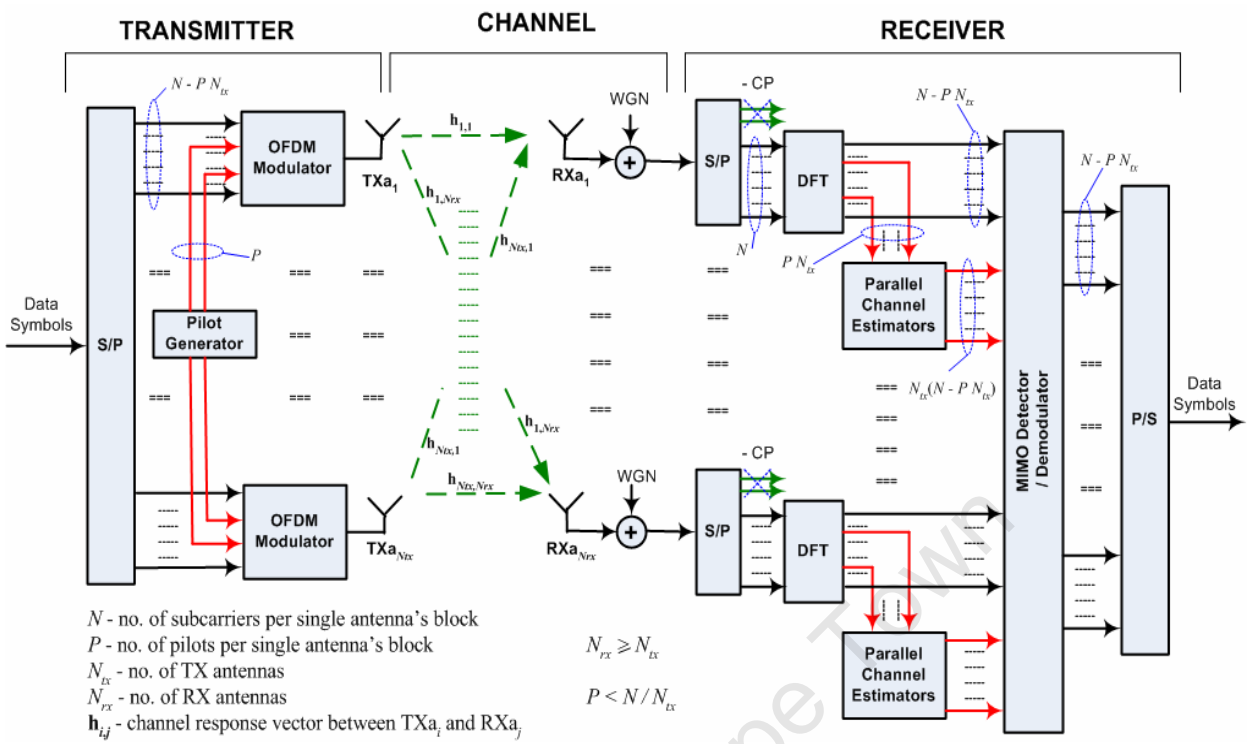


Fig.2.11. Pilot-assisted baseband SM-MIMO-OFDM system

In the transmitter, a serial binary stream is divided into N_{tx} parallel streams, each of which passes through a PSK or QAM modulator, creating a block of N complex-valued signal symbols $x_{n,m}$, $n \in [0, N-1]$ at its output, where m denotes the serial index of the block ($m \geq 0$). This block is then IDFT-transformed, prepended with the CP to eliminate IBI, converted to a serial sequence of samples and transmitted by the corresponding Tx antenna.

At the receiver side, the block at the output of each Rx antenna is separated from the CP and DFT-transformed after that. The resultant N_{rx} sets of symbols are parallel-forwarded for processing in the channel estimator and MIMO detector. Since the OFDM system can be interpreted as N parallel narrowband channels, which do not interfere with each other in frequency (Section 2.1), but are subject to SM, the received symbol on the nth subcarrier at the jth antenna within the mth block is described as [122]

$$y_{n,m}(j) = \sum_{i=0}^{N_{\text{tx}}-1} x_{n,m}(i)\, h_{n,m}(j,i) + w_{n,m}(j), \tag{2.46}$$

where $h_{n,m}(j,i)$ is the CFR gain corresponding to the (i,j)th SL (Fig.2.12), and $w_{n,m}(j)$ is the WGN sample, which represents a realisation of the CGRV with $\mathrm{E}[w_{n,m}(j)] = 0$ and $\mathrm{E}[|w_{n,m}(j)|^2] = \sigma_{\text{w}}^2$ over all j,n,m.

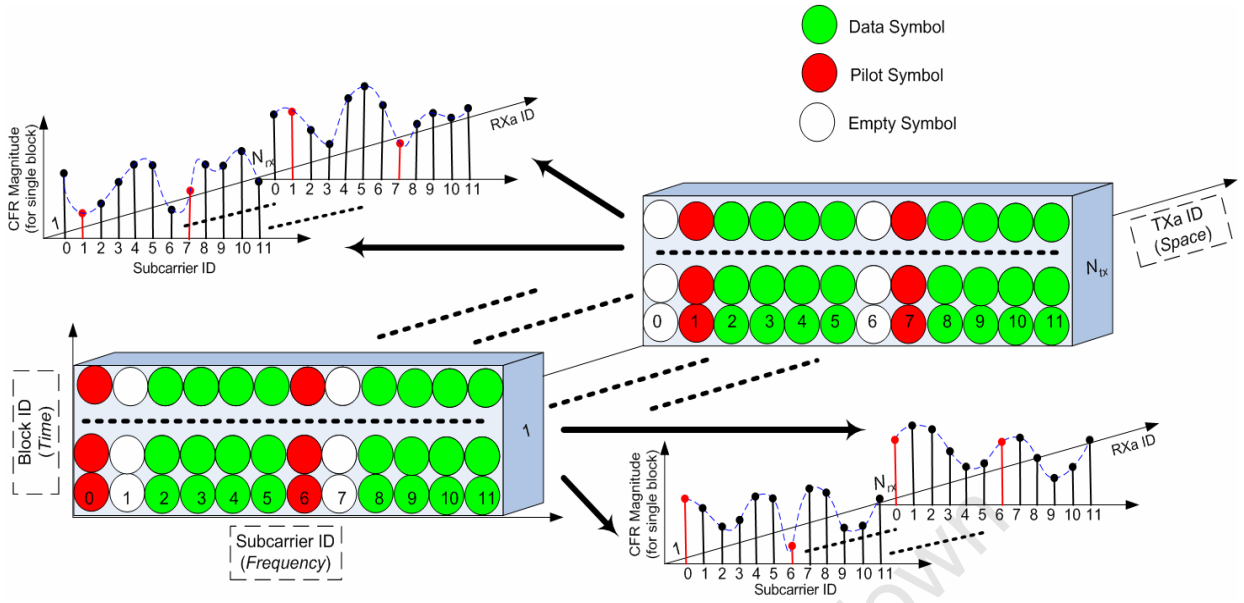


Fig.2.12. Schematic description of the transmitted signal in the pilot-assisted SM-MIMO-OFDM configuration (the system has 12 subcarriers)

2.3.3 Detection of SM data symbols

The detection algorithms are generally divided into two classes: *linear* and *non-linear*. The non-linear approaches typically offer better performance than their linear counterparts. Maximum likelihood detector (MLD) is known to be the optimal non-linear detector of the lattice-type signals [123]. However its implementation, having computational effort of order $M^{N_{tx}}$, where M is the size of the modulation constellation, is prohibitively complex for systems with many antennas and high-rate modulation schemes. Although the localised search algorithms, which perform suboptimal decoding on a selected sphere within the lattice, were proposed to reduce MLD complexity [124][125], their computational load still depends on the noise variance and is excessive for the lower SNRs. In this connection simpler linear and non-linear suboptimal detection schemes have received attention in most system prototypes of higher spatial diversity orders. In this subsection, we will investigate a linear detector that is characterised by the lowest complexity, and present its non-linear extension relying on the *decision feedback*. Note, however, that the optimal detector design is not a research objective in this thesis. Therefore for the selected system model we adopt detection algorithms, which, according to the previous research reports, are known to be efficient both from a performance and complexity standpoint.

The linear problem of detection of data symbols transmitted through the MIMO channel follows from transformation (2.46) written in the matrix notation as

$$\mathbf{y}_{n,m} = \mathbf{H}_{n,m}\mathbf{x}_{n,m} + \mathbf{w}_{n,m}, \tag{2.47}$$

where $\mathbf{x}_{n,m} = [x_{n,m}(i)]_{N_{\text{tx}}\times 1}$, $i \in [0, N_{\text{tx}} - 1]$, $n \in [0, N-1]$, $m \geq 0$ is the vector of data symbols at the input of N_{tx} Tx antennas on the *n*th subcarrier inside the *m*th block, $\mathbf{y}_{n,m} = [y_{n,m}(j)]_{N_{\text{rx}}\times 1}$, $j \in [0, N_{\text{rx}} - 1]$ is the vector of symbols at the output of N_{rx} Rx antennas, $\mathbf{w}_{n,m} = [w_{n,m}(j)]_{N_{\text{rx}}\times 1}$ is the vector of WGN samples affecting received signal, and $\mathbf{H}_{n,m} = [h_{n,m}(j,i)]_{N_{\text{rx}}\times N_{\text{tx}}}$, $i \in [0, N_{\text{tx}} - 1]$, $j \in [0, N_{\text{rx}} - 1]$ is the MIMO transform matrix.

Herein the data symbols $x_{n,m}(i)$, transmitted by different antennas, are assumed to be random, statistically independent and to have the average power σ_{d}^2, i.e. $\text{E}[\mathbf{x}_{n,m}\mathbf{x}_{n,m}^{\text{H}}] = \sigma_{\text{d}}^2 \mathbf{I}_{N_{\text{tx}}\times N_{\text{tx}}}$. ($\sigma_{\text{d}}^2$ is a fixed power for the constant modulus signals.) The noise samples are characterised as independent identically distributed (IID) CGRVs with $\text{E}[\mathbf{w}_{n,m}] = \mathbf{0}$ and $\text{E}[\mathbf{w}_{n,m}\mathbf{w}_{n,m}^{\text{H}}] = \sigma_{\text{w}}^2 \mathbf{I}_{N_{\text{rx}}\times N_{\text{rx}}}$.

2.3.3.1 Linear detector

For the known CSI, $\mathbf{H}_{n,m}$, at the receiver, the basic *zero-forcing* (ZF) detector [16] finds the estimate of transmitted symbols as a least squares (LS) solution of (2.47):

$$\hat{\mathbf{x}}_{n,m} = \mathbf{H}_{n,m}^{+}\mathbf{y}_{n,m}, \tag{2.48}$$

where $\mathbf{H}_{n,m}^{+} = \left(\mathbf{H}_{n,m}^{\text{H}}\mathbf{H}_{n,m}\right)^{-1}\mathbf{H}_{n,m}^{\text{H}}$.

Detector (2.48) is able to produce a solution only if the "tall" matrix $\mathbf{H}_{n,m}$ has a full column rank, i.e. $N_{\text{rx}} \geq N_{\text{tx}}$, that guarantees existence of the inverse of $\mathbf{H}_{n,m}^{\text{H}}\mathbf{H}_{n,m}$. It should be noted that in case of the ill-conditioned $\mathbf{H}_{n,m}^{\text{H}}\mathbf{H}_{n,m}$ ZF detector leads to dramatic amplification of the noise manifested as a considerable increase of the sum of squared errors, $\|\hat{\mathbf{x}}_{n,m} - \mathbf{x}_{n,m}\|^2$. This performance degradation is especially observed for the square systems (having $N_{\text{rx}} = N_{\text{tx}}$) and tends to grow with the number of antennas [126]. To insure existence of $\left(\mathbf{H}_{n,m}^{\text{H}}\mathbf{H}_{n,m}\right)^{-1}$ in (2.48), we can make an optimal replacement of

$$\mathbf{H}_{n,m}^{+} \approx [\mathbf{I} - (\chi^{-1}\mathbf{H}_{n,m}^{\text{H}}\mathbf{H}_{n,m} + \mathbf{I})^{-1}]\mathbf{H}_{n,m}^{+} = (\mathbf{H}_{n,m}^{\text{H}}\mathbf{H}_{n,m} + \chi\mathbf{I})^{-1}\mathbf{H}_{n,m}^{\text{H}}, \tag{2.49}$$

where $\chi \to 0$, so that $(\mathbf{H}_{n,m}^{\text{H}}\mathbf{H}_{n,m} + \chi\mathbf{I})^{-1}\mathbf{H}_{n,m}^{\text{H}} \to \mathbf{H}_{n}^{+}$, and the identity $\mathbf{I} - (\mathbf{A} + \mathbf{I})^{-1} = (\mathbf{A} + \mathbf{I})^{-1}\mathbf{A}$ [102] is used.

If $\overline{\mathbf{H}}_{n,m} = \left[\mathbf{H}_{n,m}^{\text{T}} \quad \chi^{1/2}\mathbf{I}_{N_{\text{tx}}\times N_{\text{tx}}}\right]^{\text{T}}$ and $\chi = \sigma_{\text{w}}^2/\sigma_{\text{d}}^2 = SNR_{\text{d}}^{-1}$, then substitution of the weighting matrix $\overline{\mathbf{H}}_{n,m}^{+}$ and the vector $\overline{\mathbf{y}}_{n,m} = \left[\mathbf{y}_{n,m}^{\text{T}} \quad \mathbf{0}_{N_{\text{tx}}\times 1}^{\text{T}}\right]^{\text{T}}$ into (2.48) instead of $\mathbf{H}_{n,m}^{+}$ and $\mathbf{y}_{n,m}$ yields *MMSE* detector [16][126], which is known as the optimal detector in the linear class.

The entries of $\hat{\mathbf{x}}_{n,m}$ can subsequently be demodulated (sliced on the constellation), restoring sent data. The simplest way is parallel demodulation of the symbols $\hat{\mathbf{x}}_{n,m}$ from the linear detector output. Coupling of the linear LS solution with the decision feedback from the demodulator to improve detection accuracy is known as the *V-BLAST* algorithm [127] and is presented in the next subsection.

2.3.3.2 Optimal V-BLAST detector

V-BLAST (vertical Bell Labs layered architecture for space-time communication) formulates detection as a procedure of successive cancellation of the spatial interference, starting from the most reliable decision, i.e. the symbol with the largest average signal-to-noise power ratio (SNR). Detection of this first symbol is the same as in the linear method by slicing the corresponding entry of $\hat{\mathbf{x}}_{n,m}$ (2.48) on the constellation. After that the demodulated symbol's contribution is cancelled from the received signal that is followed by zeroing the corresponding column of the MIMO transform matrix $\mathbf{H}_{n,m}$ (or $\overline{\mathbf{H}}_{n,m}$ if the MMSE smoothing is adopted). The following step determines which symbol is to undergo the linear processing next, using the updated MIMO matrix and received vector, i.e. ordering is established. For optimal performance this ordering has to be updated after each recursion. The described process is repeated until the last symbol is demodulated.

It should be noted that although the ordering update following each decision represents an optimal procedure, the original criterion to establish priorities of the detected symbols, namely average post-detection noise power [127], is suboptimal and can be replaced by a more accurate metric as proposed by Kim [128]. It has been shown that the log-likelihood ratio (LLR), exploiting both average and instantaneous noise power, yields better ordering that significantly improves performance for the case of BPSK modulation. However for QPSK and QAM performance gain is minor.

The classical V-BLAST algorithm with ZF weighting is presented in Tab.2.1. (The input parameters, corresponding to the MMSE weighting, are indicated in the brackets.) Tab.2.2 lists the *Gauss-Jordan elimination* procedure used to compute the pseudoinverse in step (a) in Tab.2.1.

Tab.2.1. Classical V-BLAST

Set $\tilde{\mathbf{y}} = \mathbf{y}_{n,m}$ ($\tilde{\mathbf{y}} = \overline{\mathbf{y}}_{n,m}$), $\tilde{\mathbf{H}} = \mathbf{H}_{n,m}$ ($\tilde{\mathbf{H}} = \overline{\mathbf{H}}_{n,m}$)

For $i = 0$ to $N_{\text{tx}} - 1$ do

(a) $\mathbf{G} := \tilde{\mathbf{H}}^{+\text{T}}$

(b) $k_i := \underset{\substack{j \in [0,\, N_{\text{tx}}-1] \\ j \neq k_l,\ l \in [0,\, i)}}{\arg\min} \left\| (\mathbf{G})_j \right\|^2$

(c) $\tilde{x}(k_i) := \text{demod}[(\mathbf{G})_{k_i}^{\text{T}}\, \tilde{\mathbf{y}}]$

(d) $\tilde{\mathbf{y}} := \tilde{\mathbf{y}} - (\tilde{\mathbf{H}})_{k_i}\, \tilde{x}(k_i)$

(e) $(\tilde{\mathbf{H}})_{k_i} := \mathbf{0}$

Return $\tilde{\mathbf{x}}_{n,m} = [\tilde{x}(0) \quad \cdots \quad \tilde{x}(N_{\text{tx}} - 1)]^{\text{T}}$

Tab.2.2. Pseudoinverse computation

Set $\mathbf{K} = \tilde{\mathbf{H}}^{\mathrm{H}}\tilde{\mathbf{H}}$, $\mathbf{L} = \mathbf{I}_{N_{\mathrm{tx}} \times N_{\mathrm{tx}}}$

Set $[\mathbf{K}]_{k_m,\, k_n} := \begin{cases} 0, & \text{if } m \neq n \\ 1, & \text{if } m = n \end{cases}$, where $m, n \in [0, i-1]$ $(i \geq 1)$

For $r = 0$ to $N_{\mathrm{tx}} - 1$ do

(a.1) $\mathbf{k} := [\mathbf{K}]_{r,r}^{-1} (\mathbf{K}^{\mathrm{T}})_r - (\mathbf{I})_r$

(a.2) $\mathbf{K} := \mathbf{K} - (\mathbf{K})_r \mathbf{k}^{\mathrm{T}}$

(a.3) $\mathbf{L} := \mathbf{L} - (\mathbf{L})_r \mathbf{k}^{\mathrm{T}}$

Return $\tilde{\mathbf{H}}^{+} = \mathbf{L}\mathbf{K}^{-1}\tilde{\mathbf{H}}^{\mathrm{H}}$, where $\mathbf{K}$ is reduced to diagonal shape

2.3.3.3 Reduced-complexity V-BLAST detector

It can be seen from Tab.2.1 and Tab.2.2 that recursive computation of the pseudoinverse represents the main complexity contribution. If the decoding order is known a priori, then V-BLAST implementation can be considerably simplified by performing the row permutation in $\tilde{\mathbf{y}}$ and the identical column permutation in $\tilde{\mathbf{H}}$ and applying the *QR decomposition* (QRD) to the resultant matrix [129]. The new form enables low-complexity successive detection without the pseudoinverse computation, starting from the symbol with the highest SNR (to be the last in the permuted vector $\tilde{\mathbf{y}}$) up to the symbol with the smallest SNR (to be the first in the permuted $\tilde{\mathbf{y}}$). *Gram-Schmidt orthogonalisation* (GSO) procedure is known as an efficient means to calculate the QR-decomposed form of a matrix (both square and "tall"). The reduced-complexity version of the detector is presented in Tab.2.3.

Note that QRD in Tab.2.3 needs ordering information. It can be obtained by computing QRD with an arbitrary ordering, followed by sorting symbols according to the column norm minimum computed for the transposed inverse of the permuted triangular matrix $\mathbf{P}$ as shown in Tab.2.4. This approach is however somewhat inefficient due to the repeated QRD computation. In Subsection 2.2.4, a constraint on channel variation has been introduced to allow for the appropriate design of the block processing system. The limited Doppler spread (2.41) implies a slow change of the channel response between two successive blocks. Hence symbol ordering for the current block can be assumed to be the same as for the previous block. Thus, QRD (Tab.2.3) is computed only once per block, followed by calculation of the ordering information (Tab.2.4), which is buffered to be used in the next block. Note also that, in contrast to the classical V-BLAST necessitating inverse computation for the matrix $\mathbf{K} = \tilde{\mathbf{H}}^{\mathrm{H}}\tilde{\mathbf{H}}$, which generally has no zero elements, the input of the ordering procedure in Tab.2.4 represents the inverse of the triangular matrix $\mathbf{P}$. In such a case the Gauss-Jordan elimination complexity is lesser as shown by an example in Appendix B.1.

Tab.2.3. Reduced-complexity V-BLAST through sorted QR decomposition

Set $\tilde{\mathbf{H}} = \mathbf{H}_{n,m}$ ($\tilde{\mathbf{H}} = \overline{\mathbf{H}}_{n,m}$), $\mathbf{Q} = \mathbf{0}_{N_{\mathrm{rx}} \times N_{\mathrm{tx}}}$ ($\mathbf{Q} = \mathbf{0}_{(N_{\mathrm{rx}}+N_{\mathrm{tx}}) \times N_{\mathrm{tx}}}$)
Set k_i , $i \in [0, N_{\mathrm{tx}} - 1]$
For $i = N_{\mathrm{tx}} - 1$ down to 0 do
(a) $(\mathbf{P})_{k_i} := \mathbf{Q}^{\mathrm{H}} (\tilde{\mathbf{H}})_{k_i}$
(b) $\mathbf{q} := (\tilde{\mathbf{H}})_{k_i} - \mathbf{Q}(\mathbf{P})_{k_i}$
(c) $[\mathbf{P}]_{k_i,k_i} := \lVert \mathbf{q} \rVert$
(d) $(\mathbf{Q})_{k_i} := [\mathbf{P}]^{-1}_{k_i,k_i} \mathbf{q}$
Return
Set $\tilde{\mathbf{y}} = \mathbf{Q}^{\mathrm{H}} \mathbf{y}_{n,m}$ ($\tilde{\mathbf{y}} = \mathbf{Q}^{\mathrm{H}} \overline{\mathbf{y}}_{n,m}$)
For $i = 0$ to $N_{\mathrm{tx}} - 1$ do
(e) $\tilde{x}(k_i) := \mathrm{demod}\left[[\mathbf{P}]^{-1}_{k_i,k_i} [\tilde{\mathbf{y}}]_{k_i} \right]$
(f) $\tilde{\mathbf{y}} := \tilde{\mathbf{y}} - (\mathbf{P})_{k_i} \tilde{x}(k_i)$
Return $\tilde{\mathbf{x}}_{n,m} = [\tilde{x}(0) \quad \cdots \quad \tilde{x}(N_{\mathrm{tx}} - 1)]^{\mathrm{T}}$

Tab.2.4. Suboptimal ordering based on single inversion of QR-decomposed MIMO transform matrix $\tilde{\mathbf{H}} = \mathbf{Q}\mathbf{P}$

Set $\mathbf{S} = \mathbf{P}^{-\mathrm{T}}$
For $i = 0$ to $N_{\mathrm{tx}} - 1$ do
$k_i := \underset{\substack{j \in [0,\ N_{\mathrm{tx}}-1] \\ j \neq k_l,\ l \in [0,\ i)}}{\arg\min} \left\lVert (\mathbf{S})_j \right\rVert^2$
Return

2.3.3.4 Computational complexity

Appendices B.2 and B.3 list several examples, which can help to assess the number of mathematical operations required by each of the presented detection algorithms. Using these examples, we generalise numerical expressions of the computational load for the system with ZF detector weighting and arbitrary number of antennas. The detection complexity is defined as the total amount of complex multiplications and additions per subcarrier per block. The resultant expressions are presented in Tab.2.5 for the general SM-MIMO configuration, and their reduced versions in case of the square system are given in Tab.2.6. Here the multiplicative and additive complexities correspond to

(B.5)-(B.6) for the linear detector, (B.7)-(B.10) for the V-BLAST with optimal ordering, (B.15)-(B.18) for the sorted QRD (SQRD) with pre-detection ordering and (B.19)-(B.22) for SQRD with post-detection ordering.

Tab.2.5. Computational complexity of different detectors

Detector	**Multiplicative complexity (C_{M})**	**Additive complexity (C_{A})**
ZF	$(N_{\mathrm{rx}}+N_{\mathrm{tx}})N_{\mathrm{tx}}^2+N_{\mathrm{rx}}N_{\mathrm{tx}}$	$(N_{\mathrm{rx}}+N_{\mathrm{tx}}-2)N_{\mathrm{tx}}^2$
ZF-VBLAST with optimal ordering	$\left(\frac{1}{3}N_{\mathrm{rx}}+\frac{1}{4}N_{\mathrm{tx}}+\frac{2}{3}\right)N_{\mathrm{tx}}^3+\frac{1}{4}(6N_{\mathrm{rx}}-3)N_{\mathrm{tx}}^2+\frac{1}{6}(19N_{\mathrm{rx}}+5)N_{\mathrm{tx}}-(2N_{\mathrm{rx}}+1)$	$\left(\frac{1}{3}N_{\mathrm{rx}}+\frac{1}{4}N_{\mathrm{tx}}-\frac{1}{6}\right)N_{\mathrm{tx}}^3+\left(N_{\mathrm{rx}}-\frac{5}{4}\right)N_{\mathrm{tx}}^2+\frac{1}{6}(16N_{\mathrm{rx}}-11)N_{\mathrm{tx}}-(2N_{\mathrm{rx}}-1)$
ZF-SQRD with pre-detection ordering	$\left(2N_{\mathrm{rx}}+\frac{1}{6}N_{\mathrm{tx}}+2\right)N_{\mathrm{tx}}^2+\left(3N_{\mathrm{rx}}+2s+\frac{5}{6}\right)N_{\mathrm{tx}}$	$\left(2N_{\mathrm{rx}}+\frac{1}{6}N_{\mathrm{tx}}+\frac{1}{2}\right)N_{\mathrm{tx}}^2+\left(N_{\mathrm{rx}}-\frac{19}{6}\right)N_{\mathrm{tx}}$
ZF-SQRD with post-detection ordering	$\left(N_{\mathrm{rx}}+\frac{1}{6}N_{\mathrm{tx}}+2\right)N_{\mathrm{tx}}^2+\left(2N_{\mathrm{rx}}+s+\frac{5}{6}\right)N_{\mathrm{tx}}$	$\left(N_{\mathrm{rx}}+\frac{1}{6}N_{\mathrm{tx}}+1\right)N_{\mathrm{tx}}^2+\left(N_{\mathrm{rx}}-\frac{8}{3}\right)N_{\mathrm{tx}}$

Tab.2.6. Computational complexity of different detectors in square antenna configuration ($N_{\mathrm{rx}}=N_{\mathrm{tx}}$)

Detector	**Multiplicative complexity (C_{M})**	**Additive complexity (C_{A})**
ZF	$2N_{\mathrm{tx}}^3+N_{\mathrm{tx}}^2$	$2N_{\mathrm{tx}}^3-2N_{\mathrm{tx}}^2$
ZF-VBLAST with optimal ordering	$\frac{7}{12}N_{\mathrm{tx}}^4+\frac{13}{6}N_{\mathrm{tx}}^3+\frac{29}{12}N_{\mathrm{tx}}^2-\frac{7}{6}N_{\mathrm{tx}}-1$	$\frac{7}{12}N_{\mathrm{tx}}^4+\frac{5}{6}N_{\mathrm{tx}}^3+\frac{17}{12}N_{\mathrm{tx}}^2-\frac{23}{6}N_{\mathrm{tx}}+1$
ZF-SQRD with pre-detection ordering	$\frac{13}{6}N_{\mathrm{tx}}^3+5N_{\mathrm{tx}}^2+\left(2s+\frac{5}{6}\right)N_{\mathrm{tx}}$	$\frac{13}{6}N_{\mathrm{tx}}^3+\frac{3}{2}N_{\mathrm{tx}}^2-\frac{19}{6}N_{\mathrm{tx}}$
ZF-SQRD with post-detection ordering	$\frac{7}{6}N_{\mathrm{tx}}^3+4N_{\mathrm{tx}}^2+\left(s+\frac{5}{6}\right)N_{\mathrm{tx}}$	$\frac{7}{6}N_{\mathrm{tx}}^3+2N_{\mathrm{tx}}^2-\frac{8}{3}N_{\mathrm{tx}}$

Using a simpler square system representation (Tab.2.6) for the comparative comments, one can see that the SQRD implementation has the same order of complexity, $O(N_{tx}^3)$, as the conventional ZF. A deeper look (Appendix B.3) reveals that ZF-QRD detector without ordering, which may be regarded as a sequential ZF method, is characterised by almost two times smaller number of mathematical operations, $C_M = N_{tx}^3 + 3N_{tx}^2 + s\,N_{tx}$, $C_A = N_{tx}^3 + \frac{3}{2}N_{tx}^2 - \frac{5}{2}N_{tx}$, than the conventional linear implementation (Tab.2.6). This circumstance justifies usage of QRD to reduce computational complexity incurred by the matrix inversion. At the same time the optimal V-BLAST detector, where the ordering is updated in every recursion, needs an order of magnitude higher number of mathematical operations ($O(N_{tx}^4)$). These observations make SQRD implementation very attractive as a powerful and cost-efficient way of improving performance of the linear detection scheme.

The ZF-SQRD method with symbol detection sorted based on the previous processed block (post-detection ordering) exhibits almost two times lesser complexity than the method where sorting and detection are achieved in the same block by repeated QRDs (pre-detection ordering). Hence the former implementation is more attractive in the scenarios of moderate CSI variation.

It should be noted that further complexity reduction is possible for SM-MIMO-OFDM systems relying on the strong correlation between the adjacent subcarriers [130], so that detection ordering could be assumed the same for a subcarrier group. However the reduced number of mathematical operations is still within the same order of magnitude and performance degradation is observed in the channels with large rms delay spread.

With regard to the MMSE weighting being used instead of ZF, it was established in Appendix B.2 that it does not change the order of complexity. Moreover, for the linear scheme ZF and MMSE are characterised by virtually the same number of mathematical operations.

2.4 Numerical performance analysis of the system with perfect CSI at the receiver

This section aims to illustrate by means of simulation examples the effects of the block-wise doubly selective channel on system performance, assuming that the receiver has perfect CSI knowledge. In Subsection 2.4.2, we study frequency-selective behaviour and impact of the channel models with different delay spreads on the *symbol error rate* (SER) of the SISO-OFDM system. Subsection 2.4.3 presents SER performance results for the SM-MIMO configuration, which help to identify the most efficient detector. Note that the results presented in this section will serve as a benchmark for the simulation performance analysis in the subsequent chapters of the thesis, which deal with a more complicated system model, incorporating channel estimator.

2.4.1 System configuration

Consider a discrete-time baseband MC system with the processing block length $N = 64$ (i.e. 64 subcarriers in the effective bandwidth *B*) and QAM modulation of the subcarriers (i.e. OFDM type). CP length is set to $N_{cp} = 7$ to accommodate CIR with a modelled length of $L = 8$ samples.

It should be noted that $N = 64$ is selected here and in the subsequent experiments as it is the smallest radix-4 FFT adopted in the standardised systems [7]. Hence it ensures faster execution of Monte Carlo simulations. For the same processing block duration higher block lengths will broaden the system bandwidth and increase sampling resolution, but CIR energy spread and Doppler variation will not alter. Hence the block length is not a vital parameter in the context of performance evaluation of the selected system model.

The modelled channel is block-wise time-variant, with the Doppler spectrum given by the Jakes model [118] and the maximum angular shift $\omega_D = 0.025\pi$, which represents the tolerable maximum according to (2.41). Two different multipath models have been adopted for simulation:

Ch.1) $K = 3$ equipowered IID components (i.e. uniform MIP) underlying bandlimited CIR (2.9), with non-sample-spaced excess delays $\tau_0 = 0$, $\tau_1 = 1.7B^{-1}$ and $\tau_2 = 4.4B^{-1}$ (illustrated in Fig.2.8);

Ch.2) L sample-spaced IID components with the exponential power decay defined in (2.14) by the factor $\alpha = 2$.

Note that Ch.1 has $\tau_{rms} = 1.81B^{-1}$, and Ch.2 is characterised by $\tau_{rms} = 0.425B^{-1}$.

The modelled MIMO channel properties are in full accordance with the assumptions made in Subsection 2.3.1.

2.4.2 Symbol error rate in SISO transmission mode

Doubly selective behaviour of the channel is clearly visible in Fig.2.13, where CFR magnitude corresponding to Ch.1 is plotted for a sequence of blocks. One can see that some subcarriers may be subject to strong attenuation in a few blocks in the plotted sequence, while for the other blocks their gains are sufficient to guarantee reliable symbol transmission through the channel.

In the example studied here, OFDM subcarriers are modulated by QPSK and 16QAM. The elementary single-tap ZF equaliser performs CFR compensation at the receiver.

Fig.2.14 shows that SER of the QPSK-modulated system is about 8.5dB lower than SER of the system with 16QAM. Although CIR length is the same for both Ch.1 and Ch.2, there is a growing performance difference at higher SNRs at the receiver input. Thus, the plotted results confirm the major dependence of system performance on the rms delay spread of the channel, which has been theoretically stated in Subsection 2.2.2. Fig.2.14 shows that SER is higher for Ch.1, which has smaller number of underlying multipath components than Ch.2, but at the same time a more dispersive profile.

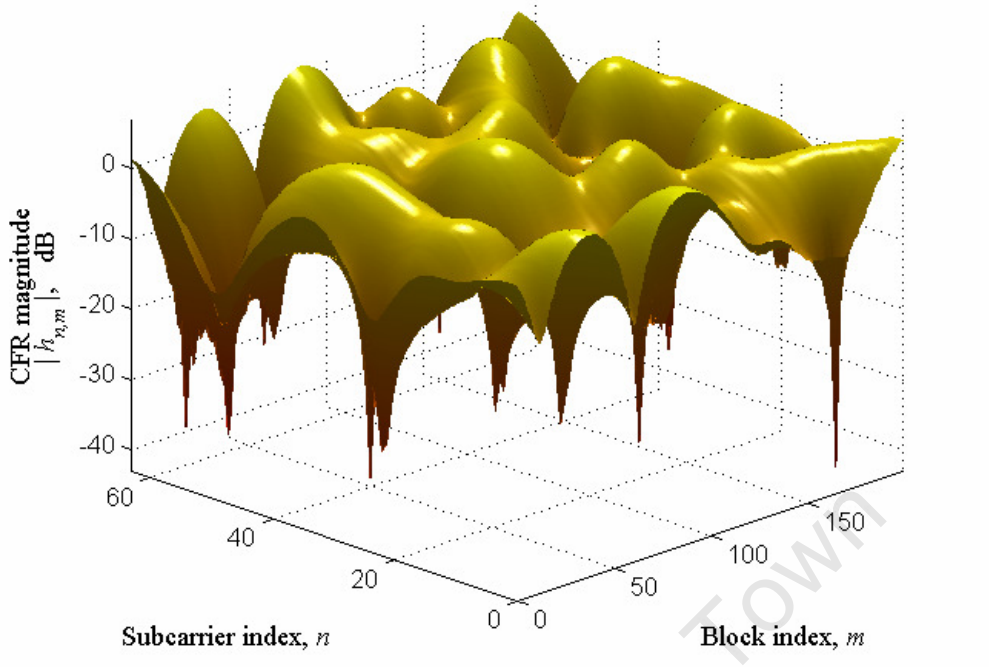


Fig.2.13. Example of CFR magnitude of the doubly selective channel

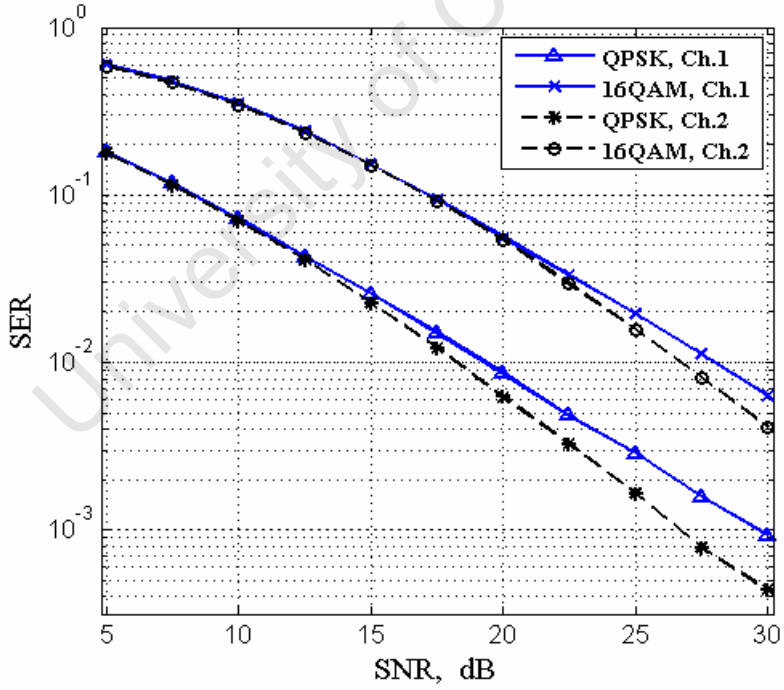


Fig.2.14. SER of the SISO-OFDM system with perfect CSI at the receiver

2.4.3 Symbol error rate in MIMO transmission mode

The SM-MIMO configuration is tested in two modes to demonstrate receive diversity impact on performance: with 4 Tx and 4 Rx antennas (4x4) and with 4 Tx and 6 Rx antennas (4x6). Selected channel model is Ch.1. The experiment compares detectors of various complexity levels, described in Subsection 2.3.3.

The square system (4x4) with the linear ZF detector is seen (Fig.2.15) to have about 5.5dB-6dB worse performance than SISO-OFDM (1x1). As explained in Subsection 2.3.3, this is caused by ill-conditioned realisations of the MIMO transform matrix. Adoption of MMSE weighting instead of ZF considerably reduces the loss (< 2dB). At higher SNRs this gap diminishes.

Increase of the receive diversity by introduction of 6 antennas instead of the minimum 4 greatly improves SER at higher SNRs. Note that the difference between ZF and MMSE becomes quite small as the additional rows in the MIMO transform matrix guarantee that it has full column rank. Hence ZF weighting in the detector is suitable for higher diversity orders and will be used in the subsequent analysis of the 4x6 configuration. Furthermore, the use of ZF detector does not imply the necessity of noise power tracking, inherent to the MMSE design.

One can also see from Fig.2.15 that using the V-BLAST detector leads to the large performance gain at higher SNRs in contrast to the linear detector (>5dB for SER 10^{-4}). The smaller SER difference at lower SNRs can be explained by propagation of the V-BLAST decisions errors, occurring in the starting recursion steps.

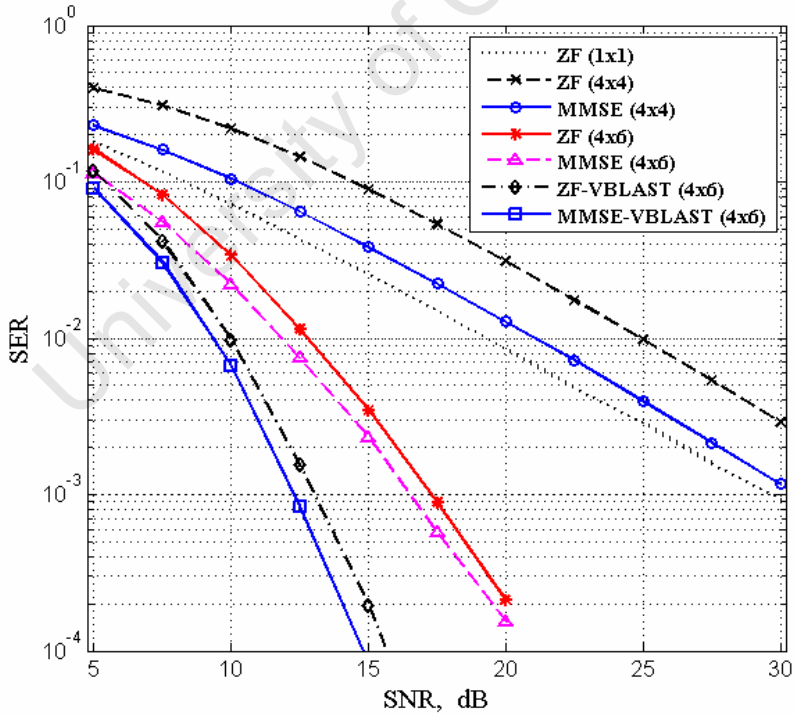


Fig.2.15. SER of the QPSK-modulated SM-MIMO-OFDM system with perfect CSI at the receiver and different detectors

SER of the system with the low-complexity V-BLAST alternative implemented as the SQRD algorithm (Tab.2.3) with post-detection ordering is shown in Fig.2.16 in comparison with the classical V-BLAST detector (Tab.2.1). Despite the suboptimal ordering, SQRD performance degradation with regard to V-BLAST is very negligible and is visible only at very high SNRs. A channel with slower variation than the simulated Doppler spread, equal to the tolerable maximum (2.41), would lead to even better results, as then the ordering is indeed the same for successive blocks.

It is interesting to note that for the case of 16QAM modulation SQRD detector with MMSE weighting exhibits almost the same performance as with ZF weighting. This is in contrast to the QPSK case, for which 0.6dB-0.7dB difference is observed between the corresponding curves.

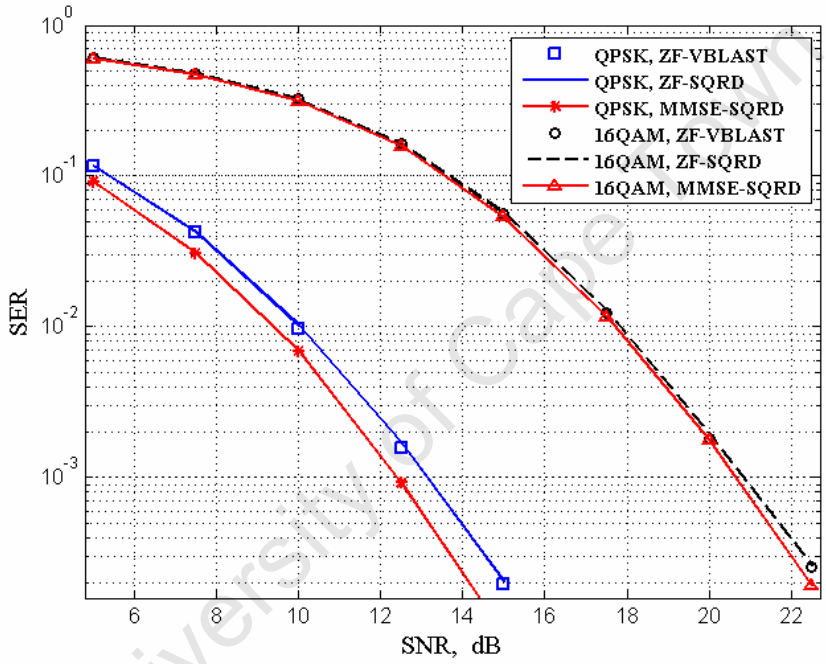


Fig.2.16. Comparative performance of the optimal and suboptimal decision feedback detection algorithms for the case of 4x6 system

2.5 Summary

MIMO doubly selective wireless channel represents a complicated theoretical concept, which has been explicitly addressed throughout this chapter by means of the LTV model development and appropriate statistical characterisation. It has been established that the MC block transmissions are possible in such type of channels under certain constraints on the time variation and multipath dispersion. Adding CP to the transmitted block eliminates IBI

and creates conditions for application of linear processing techniques to the received signal, facilitating detection and, as will be considered in Chapter 3, channel estimation.

In this thesis, the MIMO-OFDM system with spatial multiplexing is investigated. Therefore an efficient detector is a crucial design requirement. To meet this requirement, several detectors have been analysed from both the complexity and performance standpoint. It has been found that in the channels with the aforementioned time variation constraint, a low-complexity decision-feedback SQRD detector guarantees SER close to its optimal counterpart.

The system model developed so far lacks only realistic CSI acquisition that will be addressed in the subsequent material as a central research problem.

3 Low-Complexity Pilot-Assisted OFDM Channel Estimators

The spectrally efficient OFDM system model, introduced in Chapter 2, assumed perfect channel state information (CSI) at the receiver. It allowed implementation of coherent detection algorithms, recovering sent data with minimum errors in a high signal-to-noise power ratio (SNR) regime. However, in real wireless communication systems, adaptation of the processing architecture of the receiver to the actual channel conditions is challenging. In contrast to the wired communication links, the radio fading channel characteristics strongly depend on the physical features of the propagation environment (indoor, outdoor, etc.). Hence, for each new transceiver location the channel response would be represented by a completely random function. The problem is aggravated by the time variation of the channel response that can become significant if the transmitter and receiver are mobile with regard to each other. These two circumstances imply the necessity to perform CSI acquisition on a regular basis and re-train the channel estimator at the receiver if the channel statistics change.

It was mentioned in Chapter 1 that for the multicarrier systems, loading pilot symbols on the subset of subcarriers provides the best means of tracking the time-varying channel. We begin this chapter with the formulation of pilot-aided channel response measurements as a key idea towards partial CSI acquisition (Section 3.1). The subsequent analyses will be dedicated to the optimal and suboptimal linear processing methods, aiming to clean these measurements from noise and to restore entire CSI required for the detector. In Section 3.2, the simplest constrained least squares (CLS) estimator is presented as a starting point to the derivation of the more advanced algorithms, which are explicitly described in Sections 3.3 and 3.4, with the emphasis on computational complexity and analytical performance. The chapter is concluded by the comparative complexity analysis in Section 3.5 and numerical examples in Section 3.6.

3.1 Pilot-assisted channel estimation as a linear problem

Placement of the *pilot symbols* onto subcarriers inside the OFDM block is schematically depicted in Fig.2.12. Such a pilot structure is known as "*independent*" and is characterised by a dedicated pilot subcarrier set assigned to each Tx antenna, which do not overlap with the data-modulated subcarriers and with the pilot sets of other antennas.

The independent pilot pattern has been found optimal in the research literature [100], and has also been adopted in the optional MIMO mode, specified in the commercially successful IEEE 802.16 standard [20].

Note that for successful decoupling of channel estimation at the spatial layers (SLs), associated with different Tx antennas, it is necessary to place zero-valued *virtual symbols* (VSs) at the subcarrier, where the pilot symbol is transmitted by the *i*th Tx antenna (Fig.2.12). These VSs have to be loaded on the corresponding subcarrier by the other $N_{\text{tx}}-1$ antennas in order to exclude spatial multiplexing (SM) interference and thus to enable low-complexity channel estimation for all $(i,j), \forall j \in [0, N_{\text{rx}}-1]$ SLs.

Let $\mathbf{x}_m^{\mathbf{p}}(i) = [x_{(p_k+i) \bmod N, m}(i)]_{P\times 1}$, $k \in [0, P-1]$ denote the vector of the pilot symbols transmitted by the *i*th Tx antenna, inside the *m*th block, at the subcarriers indexed $(p_k + i) \bmod N$, $p_k \in [0, N-1]$, such that $[p_{(k+1) \bmod P} - p_k + N] \bmod N \geq N_{\text{tx}}$ for $\forall k \in [0, P-1]$, where the selected pilot pattern $\{p_k\}$ is cyclically shifted on different Tx antennas to perform independent channel estimation at various SLs. The positions of VSs loaded on the *i*th Tx antenna can be specified as $q_{k,b}(i) = (p_k + b) \bmod N$, where $b \in [0, N_{\text{tx}}-1]$, $b \neq i$. (For other parameters we use the same notations as in Chapter 2.)

Under the aforementioned pilot sequence separation property, we describe the vector of channel-transformed pilot symbols received by the *j*th Rx antenna from the *i*th Tx antenna inside the *m*th block in analogy with the linear equation (2.4) as

$$\mathbf{y}_m^{\mathbf{p}}(j,i) = [y_{(p_k+i) \bmod N, m}(j)]_{P\times 1} = \mathbf{X}^{\mathbf{p}}(i)\mathbf{C}(i)\mathbf{h}_m(j,i) + \mathbf{w}_m^{\mathbf{p}}(j,i), \tag{3.1}$$

where $\mathbf{w}_m^{\mathbf{p}}(j,i) = [w_{(p_k+i) \bmod N, m}(j)]_{P\times 1}$ is a stack of P WGN samples, affecting the output of the *j*th Rx antenna ($\mathrm{E}[\mathbf{w}_m^{\mathbf{p}}(j,i)\mathbf{w}_m^{\mathbf{p}}(j,i)^{\mathrm{H}}] = \sigma_{\mathrm{w}}^2 \mathbf{I}_{P\times P}$), $\mathbf{h}_m(j,i) = [h_{n,m}(j,i)]_{N\times 1}$, $n \in [0, N-1]$ is the *channel frequency response* (CFR) vector, the diagonal matrix $\mathbf{X}^{\mathbf{p}}(i) = \mathrm{dg}\left(\mathbf{x}^{\mathbf{p}}(i)\right)$ contains reference values of P *equipowered* pilot symbols in the *i*th Tx antenna's spectrum (we assume that the pilot sequence is the same for $\forall m$), and $\mathbf{C}(i)$ is the $P \times N$ size selection matrix that is needed to extract samples of the layer-specific CFR $\mathbf{h}_m(j,i)$, corresponding to the pilot subcarriers. The elements of $\mathbf{C}(i)$ are expressed as $[\mathbf{C}(i)]_{k,n} = \begin{cases} 1, & \text{if } n = (p_k + i) \bmod N \\ 0, & \text{otherwise} \end{cases}$.

As an essential design requirement, expression (3.1) should guarantee perfect CFR acquisition by the receiver in the absence of noise, i.e. when $\mathbf{w}_m^{\mathbf{p}}(j,i) = \mathbf{0}_{P\times 1}$. Since CFR, $\mathbf{h}_m(j,i)$, represents a vector of length $N > P$, the direct restoration of $\mathbf{h}_m(j,i)$ based on $\mathbf{y}_m^{\mathbf{p}}(j,i)$ is impossible. In Section 2.1, the channel was characterised by the *channel impulse response* (CIR) realisation $\underline{\mathbf{h}}_m(j,i)$ of finite length $L < N$ that is linked to the corresponding CFR through the linear combination

$$\mathbf{h}_m(j,i) = \mathbf{F}\,\mathbf{B}\,\underline{\mathbf{h}}_m(j,i). \tag{3.2}$$

If $\underline{\mathbf{h}}_m(j,i)$ is unknown, then (3.2) represents an overdetermined system of linear equations. Its solution can be found using any L samples of $\mathbf{h}_m(j,i)$. In our model, P pilot symbols are used to restore CFR in the entire band, hence it is required that $P \geq L$.

Substituting (3.2) into (3.1), where $\mathbf{w}_m^{\mathrm{p}}(j,i) = \mathbf{0}$, yields $\mathbf{y}_m^{\mathrm{p}}(j,i) = \mathbf{X}^{\mathrm{p}}(i)\mathbf{C}(i)\mathbf{F}\mathbf{B}\underline{\mathbf{h}}_m(j,i)$. From here the unknown CIR vector is found as

$$\begin{aligned}\underline{\mathbf{h}}_m(j,i) &= (\mathbf{X}^{\mathrm{p}}(i)\mathbf{C}(i)\mathbf{F}\mathbf{B})^{+}\mathbf{y}_m^{\mathrm{p}}(j,i)\\ &= (\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}(i)^{\mathrm{H}}\mathbf{X}^{\mathrm{p}}(i)^{\mathrm{H}}\mathbf{X}^{\mathrm{p}}(i)\mathbf{C}(i)\mathbf{F}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}(i)^{\mathrm{H}}\mathbf{X}^{\mathrm{p}}(i)^{\mathrm{H}}\mathbf{y}_m^{\mathrm{p}}(j,i)\\ &= \sigma_{\mathrm{p}}^{-2}(\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}(i)^{\mathrm{H}}\mathbf{C}(i)\mathbf{F}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}(i)^{\mathrm{H}}\mathbf{X}^{\mathrm{p}}(i)^{\mathrm{H}}\mathbf{y}_m^{\mathrm{p}}(j,i)\end{aligned}, \quad (3.3)$$

where σ_{p}^2 is the power of the pilot symbol.

After $\underline{\mathbf{h}}_m(j,i)$ is found, the corresponding $\mathbf{h}_m(j,i)$ function can be restored in full by (3.2). Thus, it is important to note here that ideal CFR acquisition is possible in the absence of noise when CIR has finite length, which is lesser than the block length. The combination of formulas (3.3) and (3.2) is known as the DFT *interpolator*. Its principle is based on the postulate that interpolation is a linear operation, which represents a product of the observation vector, containing function knots, with a "tall" matrix. One can see from the corresponding equations that the DFT interpolator can be described as a cascade of three linear processing stages:

1) IDFT of zero-interleaved observation vector, $\mathbf{C}(i)^{\mathrm{H}}\mathbf{X}^{\mathrm{p}}(i)^{\mathrm{H}}\mathbf{y}_m^{\mathrm{p}}(j,i)$, with output truncated to the size L;

2) product with the matrix $(\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}(i)^{\mathrm{H}}\mathbf{C}(i)\mathbf{F}\mathbf{B})^{-1}$, yielding CIR vector $\underline{\mathbf{h}}_m(j,i)$;

3) DFT of zero-padded (up to the size *N*) $\underline{\mathbf{h}}_m(j,i)$.

In a special case of $P = L$, $\mathbf{h}_m(j,i) = \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}(\mathbf{C}(i)\mathbf{F}\mathbf{B})^{-1}\mathbf{X}^{\mathrm{p}}(i)^{\mathrm{H}}\mathbf{y}_m^{\mathrm{p}}(j,i)$, i.e. interpolation includes only two stages:

1) product with the $L \times L$ matrix $(\mathbf{C}(i)\mathbf{F}\mathbf{B})^{-1}$;

2) *N*-point DFT.

It is important to keep in mind the limitation on the number of pilots, imposed by the VS insertion and the anticipated order of the channel model:

$$L \leq P \leq N/N_{\mathrm{tx}}, \quad (3.4)$$

where the largest value $P = N/N_{\mathrm{tx}}$ corresponds to the case when the entire bandwidth is utilised for training (no data is transmitted within the *m*th block).

The presence of unknown noise in (3.1) imparts a statistical nature to the problem, and motivates a search of an optimal solution of the observation equation set, (3.1) and (3.2), that would be less affected by the noise term $\mathbf{w}_m^{\mathrm{p}}(j,i)$. Linear channel estimation algorithms produce CFR estimate $\hat{\mathbf{h}}_m(j,i)$ by means of a transformation

$$\hat{\mathbf{h}}_m(j,i) = \mathbf{M}(i)\mathbf{y}_m^{\mathrm{p}}(j,i), \quad (3.5)$$

where $\mathbf{M}(i)$ represents some filtering matrix. In the subsequent sections, various complexity algorithms computing $\mathbf{M}(i)$ will be described.

3.2 Constrained LS intrablock estimator

In Section 2.2, analysis of the channel model has outlined a number of deterministic and statistical properties, underlying the channel response process, namely finiteness of CIR, correlation in the frequency domain (intrablock correlation) and correlation in the time domain (interblock correlation). If the receiver has a priori or acquired information on these properties, it could be used in the channel response identification. The LS estimator, described in the subsequent subsection, relies on the CIR finiteness assumption to denoise the estimate in the excess delay domain.

3.2.1 Design and implementation

The LS problem of obtaining an estimate of the CIR $\underline{\mathbf{h}} = [\underline{h}_0 \quad \cdots \quad \underline{h}_{L-1}]^{\mathrm{T}}$ based on the pilot-assisted measurements is formulated as minimisation of the quadratic difference (sum of square errors) between the received pilot symbols $\mathbf{y}^{\mathrm{p}}$ and the reference pilot values $\mathbf{x}^{\mathrm{p}} = [x_{P0} \quad \cdots \quad x_{PP-1}]^{\mathrm{T}}$, affected by the assumed CFR model $\mathbf{h}$, that can be written using (3.1) and (3.2) as

$$J(\underline{\mathbf{h}}) = \mathbf{w}^{\mathrm{p}\,\mathrm{H}}\mathbf{w}^{\mathrm{p}} = \left(\mathbf{y}^{\mathrm{p}} - \mathbf{X}^{\mathrm{p}}\mathbf{C}\mathbf{h}\right)^{\mathrm{H}}\left(\mathbf{y}^{\mathrm{p}} - \mathbf{X}^{\mathrm{p}}\mathbf{C}\mathbf{h}\right) = \left(\mathbf{y}^{\mathrm{p}} - \mathbf{X}^{\mathrm{p}}\mathbf{C}\mathbf{F}\mathbf{B}\underline{\mathbf{h}}\right)^{\mathrm{H}}\left(\mathbf{y}^{\mathrm{p}} - \mathbf{X}^{\mathrm{p}}\mathbf{C}\mathbf{F}\mathbf{B}\underline{\mathbf{h}}\right), \tag{3.6}$$

where we have omitted block index *m* and antenna indices *i* and *j* for brevity as the processing for each block and SL is done separately from the others.

It is shown in Appendix C.1 that minimisation of (3.6) with regard to $\underline{\mathbf{h}}$ produces the *constrained LS* (CLS) CIR solution. Here the term "constrained" indicates the CIR finiteness assumption (first *L* samples only), which may not be true in general, e.g., for the non-sample-spaced channels, which are characterised by energy leakage due to the band limitation. However, this leakage can be regarded as negligible. Thus, the estimate is defined as

$$\underline{\tilde{\mathbf{h}}} = \sigma_{\mathrm{p}}^{-2}\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathrm{p}\,\mathrm{H}}\mathbf{y}^{\mathrm{p}}, \tag{3.7}$$

where the inverted Hermitean matrix $\mathbf{S} = \left(\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\mathbf{F}\mathbf{B}\right)^{-1}$ exists due to the condition $\mathrm{rk}(\mathbf{S}) = \mathrm{rk}(\mathbf{C}\mathbf{F}\mathbf{B}) = \mathrm{rk}(\mathbf{F}\mathbf{B}) = L$, which holds if and only if $\mathrm{rk}(\mathbf{F}\mathbf{B}) \le \mathrm{rk}(\mathbf{C}) = P$ [131]. However, a caution should be exercised as for certain pilot arrangements $P^{-L}\det(\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\mathbf{F}\mathbf{B})$ may be close to zero and hence $\mathbf{S}^{-1}$ is ill-conditioned with respect to inversion. One should also note that $\mathbf{S}$ is doubly symmetric as it is the inverse of the

doubly symmetric matrix $\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\mathbf{F}\mathbf{B}$ [103]. This follows from $\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\mathbf{F}$ being a circulant matrix, truncation of which from right ($\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\mathbf{F}\mathbf{B}$) and bottom ($\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\mathbf{F}$) results in a doubly symmetric matrix.

CFR estimate represents a DFT of $\underline{\tilde{\mathbf{h}}}$, i.e., similar to (3.2),

$$\tilde{\mathbf{h}} = \mathbf{F}\mathbf{B}\underline{\tilde{\mathbf{h}}} = \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\mathrm{H}}\mathbf{y}^{\mathbf{p}}. \tag{3.8}$$

When the pilot subcarriers are *equispaced*, i.e. $C_{k,n} = \begin{cases} 1, \text{ if } n = p_k = kN/P + \nu \\ 0, \text{ otherwise} \end{cases}$, where $N \bmod P = 0$, $k \in [0, P-1]$, and $0 \le \nu \le N/P - 1$, it can be shown (Appendix C.2) that $\mathbf{S} = P^{-1}\mathbf{I}$ and $\underline{\tilde{\mathbf{h}}} = \sigma_{\mathrm{p}}^{-2} P^{-1}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\mathrm{H}}\mathbf{y}^{\mathbf{p}}$.

It is interesting to point out that the CLS estimator (3.7)-(3.8) has exactly the same form as the DFT interpolator (3.2)-(3.3), though it is derived based on the noise reduction criterion, whereas the DFT interpolator does not take noise into account. The observed similarity implies that the overall procedure of the CLS algorithm (3.8) is executed in the following three-stage order:

1) zero-interleaved vector with the pilot measurement samples, $\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\mathrm{H}}\mathbf{y}^{\mathbf{p}}$, is translated to the time domain by the inverse discrete Fourier transform (IDFT);
2) the IDFT output is shrunk down to the first L elements and weighted by the matrix $\mathbf{S}$ of small ($L \times L$) dimensions;
3) the resultant vector is padded with $N - L$ zeros in the tail and converted back to the frequency domain through the DFT procedure.

The use of the DFT/IDFT pair, which enables reduction of the unknown parameter space from *N* to *L*, implies classification of the presented CLS estimator as the transform-domain processing algorithm. The biggest advantage of such techniques is low computational complexity that is achieved with no performance degradation with regard to the linear methods, performing processing entirely in the frequency domain [43][58], if the CIR finiteness assumption is satisfied. Implementation of the CLS channel estimator is shown schematically in Fig.3.1.

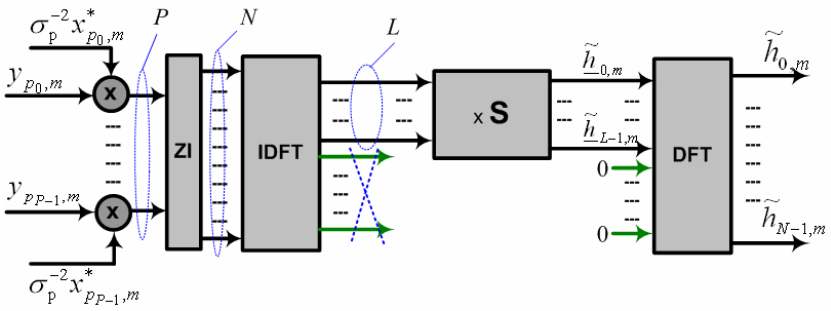


Fig.3.1. CLS estimator

3.2.2 Computational complexity

To assess computational complexity of the channel estimator, we will henceforth determine the number of *complex multiplications* and *complex additions* required to obtain CFR estimate on the interval of one OFDM block.

The time-frequency interpretation, based on the DFT/IDFT pair, significantly diminishes complexity of the CLS algorithm. If N is a power of 4, DFT (and IDFT, which is mathematically equivalent to DFT) can be computed by the radix-4 *fast Fourier transform* (FFT) algorithm, which forms the base of most contemporary OFDM systems [7]-[20] and requires $3N/8\log_2 N$ multiplies and $3N/2\log_2 N$ additions [132]. As the transform-domain CLS estimator (Fig.3.1) includes DFT with only a partial input and IDFT with partial input and output, a significant complexity reduction is achieved by exploiting the pruning FFT algorithms [133]. As shown by Khan et al. [134], in the special case of the equispaced FFT input elements, the pruning approach allows to lower the number of complex multiplications down to $N/8 + P/3\log_2 P$. In the subsequent analysis, we let the cumulative number of mathematical operations, associated with pre- and post-filtering of the L-element CIR vector, be denoted as $C_{\mathrm{M}}^{\mathrm{FFT}}$, which takes into account DFT and IDFT stage, as well as P products of y_{p_k} and $\sigma_{\mathrm{p}}^{-2}x_{p_k}^{*}$, to compute the IDFT input (Fig.3.1). Correspondingly, the cumulative number of additions for the DFT and IDFT engines will be denoted as $C_{\mathrm{A}}^{\mathrm{FFT}}$. The linear estimators developed in the current chapter all belong to the transform-domain class and will be shown to share the same pre- and post-filtering stages as the CLS estimator. Hence implementation of the FFT/IFFT part of the estimator is irrelevant from the comparative analysis standpoint.

In a general case, product of the IFFT output vector with the Hermitean positive-definite weighting matrix $\mathbf{S}$ requires L^2 multiplications and $L(L-1)$ additions. For the equispaced pilot arrangement $\mathbf{S} = P^{-1}\mathbf{I}$, hence the scaling factor P^{-1} can be included on the pre-IFFT stage, i.e. replacing $\sigma_{\mathrm{p}}^{-2}x_{p_k}^{*}$ with $\sigma_{\mathrm{p}}^{-2}P^{-1}x_{p_k}^{*}$, and the filtering stage is represented by trivial denoising of the CIR tail part. Thus, the total amount of mathematical operations (multiplies and additions) required by the CLS estimator is

$$C_{\mathrm{M}}^{\mathrm{cls}} = C_{\mathrm{M}}^{\mathrm{FFT}} + L^2, \tag{3.9}$$

$$C_{\mathrm{A}}^{\mathrm{cls}} = C_{\mathrm{A}}^{\mathrm{FFT}} + L^2 - L \tag{3.10}$$

for the case of non-equispaced pilot subcarriers, and just $C_{\mathrm{M}}^{\mathrm{cls}} = C_{\mathrm{M}}^{\mathrm{FFT}}$, $C_{\mathrm{A}}^{\mathrm{cls}} = C_{\mathrm{A}}^{\mathrm{FFT}}$ for the equispaced pilots (also known as *comb* pilot pattern [37]).

3.2.3 MSE analysis

Denote the vector of CFR estimation errors at the output of the CLS estimator (3.8) as $\tilde{\boldsymbol{\varepsilon}} = \tilde{\mathbf{h}} - \mathbf{h}$. The use of (3.7), where substitution $\mathbf{y}^{\mathbf{p}} = \mathbf{X}^{\mathbf{p}}\mathbf{C}\mathbf{F}\mathbf{B}\underline{\mathbf{h}} + \mathbf{w}^{\mathbf{p}}$ is made according to the definitions (3.1)-(3.2), yields

$$\begin{aligned}\tilde{\boldsymbol{\varepsilon}} &= \mathbf{F}\mathbf{B}(\tilde{\underline{\mathbf{h}}} - \underline{\mathbf{h}}) \\ &= \mathbf{F}\mathbf{B}(\sigma_{\mathrm{p}}^{-2}\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathrm{p}\,\mathrm{H}}\mathbf{y}^{\mathrm{p}} - \underline{\mathbf{h}}) \\ &= \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathrm{p}\,\mathrm{H}}\mathbf{w}^{\mathrm{p}} \\ &= \mathbf{F}\mathbf{B}\tilde{\underline{\boldsymbol{\varepsilon}}}\end{aligned}, \tag{3.11}$$

where $\tilde{\underline{\boldsymbol{\varepsilon}}}$ represents CIR estimation error vector. The corresponding CFR error correlation matrix is then found as follows:

$$\begin{aligned}\mathbf{R}_{\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}} &= \mathrm{E}[\tilde{\boldsymbol{\varepsilon}}\,\tilde{\boldsymbol{\varepsilon}}^{\mathrm{H}}] \\ &= \mathbf{F}\mathbf{B}\mathbf{R}_{\tilde{\underline{\boldsymbol{\varepsilon}}}\tilde{\underline{\boldsymbol{\varepsilon}}}}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}} \\ &= \sigma_{\mathrm{p}}^{-4}\mathbf{F}\mathbf{B}\,\mathrm{E}[\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathrm{p}\,\mathrm{H}}\mathbf{w}^{\mathrm{p}}\mathbf{w}^{\mathrm{p}\,\mathrm{H}}\mathbf{X}^{\mathrm{p}}\mathbf{C}\mathbf{F}\mathbf{B}\mathbf{S}]\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}} \\ &= \sigma_{\mathrm{w}}^{2}\sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\end{aligned}, \tag{3.12}$$

where $\mathbf{R}_{\tilde{\underline{\boldsymbol{\varepsilon}}}\tilde{\underline{\boldsymbol{\varepsilon}}}} = SNR_{\mathrm{p}}^{-1}\mathbf{S}$ is the correlation matrix of the CIR estimation error, and $SNR_{\mathrm{p}} = \sigma_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^{2}$ is the signal-to-noise power ratio at the pilot subcarriers.

Hence the variance of the estimation error is found as

$$\sigma_{\tilde{\varepsilon}}^{2} = N^{-1}\mathrm{tr}(\mathbf{R}_{\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}}) = \mathrm{tr}(\mathbf{R}_{\tilde{\underline{\boldsymbol{\varepsilon}}}\tilde{\underline{\boldsymbol{\varepsilon}}}}) = (s\,SNR_{\mathrm{p}})^{-1}, \tag{3.13}$$

where

$$s = 1/\mathrm{tr}(\mathbf{S}). \tag{3.14}$$

For the equispaced arrangement of the pilot subcarriers $\mathbf{S} = P^{-1}\mathbf{I}$ and $s = PL^{-1}$, so that (3.13) reduces to

$$\sigma_{\tilde{\varepsilon},\mathrm{es}}^{2} = SNR_{\mathrm{p}}^{-1}P^{-1}L. \tag{3.15}$$

Since the channel estimation errors (3.11) form a vector of complex Gaussian random variables (CGRVs) with zero mean, error variance notation is equivalent to the mean square error (MSE) notation (analogously, the covariance notation is equivalent to the correlation notation). In the sequel we will use both terms interchangeably.

It should be noted that it is possible to enhance performance of the CLS estimator, if the multipath channel is sample-spaced and the number of paths is lesser than the CIR length. In our related work [42], not included in this thesis, the low-complexity modified order-recursive least squares (MORLS) algorithm has been proposed. It constructs CIR estimate successively and allows filtering noise for each CIR sample individually. One of the useful properties inherent to the MORLS estimator is estimation of the CIR length, providing the means for prediction and adaptive assignment of the cyclic prefix (CP) length that would reduce transmission redundancy for the less dispersive channels.

3.2.4 Extension to MIMO

Under the homogeneity assumption of the spatial channel, the detector is able to restore transmitted data symbols only if CSI from at least N_{tx}^{2} SLs is available (refer to Section 2.3). It has been shown by the simulation

examples (Subsection 2.4.3) that for the systems with receive diversity ($N_{rx} > N_{tx}$), CSI from the additional $N_{tx}(N_{rx} - N_{tx})$ SLs can significantly improve detection performance. However, if the number of Tx and Rx antennas is large (four or more), the receiver complexity is greatly influenced by the channel estimation algorithm, due to the necessity to implement $N_{tx}N_{rx}$ estimators retrieving CSI from all SLs.

From the complexity standpoint, the CLS estimator, especially based on the equispaced pilot arrangement, provides an efficient means of channel identification for MIMO systems of large dimensions. In analogy with (3.3), CFR for the (i, j) th SL is acquired by computing

$$\tilde{\mathbf{h}}_m(j,i) = \sigma_p^{-2}\mathbf{F}\mathbf{B}\mathbf{S}(i)\mathbf{B}^H\mathbf{F}^H\mathbf{C}(i)^H\mathbf{X}^p(i)^H\mathbf{y}_m^p(j,i), \tag{3.16}$$

where $\mathbf{S}(i) = (\mathbf{B}^H\mathbf{F}^H\mathbf{C}(i)^H\mathbf{C}(i)\mathbf{F}\mathbf{B})^{-1}$ depends on the Tx antenna. In the special case of the equispaced pilot subcarriers, $\mathbf{S}(i) = \mathbf{S}(0) = P^{-1}\mathbf{I}$, i.e. the filtering part of the estimator is the same for all SLs, which greatly reduces complexity of the MIMO system configuration.

3.3 Constrained MMSE intrablock estimator

The CLS estimator described in the previous section relies only on the finiteness of the channel impulse response, which may be regarded as a deterministic property. Characterisation of the channel model in Section 2.2 revealed a number of second-order statistical parameters, namely CIR correlation and noise variance, which may potentially be used to improve estimation accuracy.

In this section, we derive the constrained minimum mean square error (CMMSE) intrablock estimator. It makes use of second-order channel statistics, which can be updated recursively. Although the CMMSE solution is formulated through a completely independent linear problem, it will be shown that this estimator has the same transform-domain structure as CLS, where the filtering part is supplemented with the smoothing module.

3.3.1 Design and implementation

Pilot-assisted linear MMSE estimator is constructed in the form of

$$\hat{\mathbf{h}}^{mmse} = \mathbf{Q}\mathbf{y}^p, \tag{3.17}$$

where the $N \times P$ size weighting matrix $\mathbf{Q}$ is selected to minimise MSE between the CFR estimate $\hat{\mathbf{h}}^{mmse}$ and the assumed CFR model $\mathbf{h}$, described by the correlation matrix $\mathbf{R}_{\mathbf{hh}} = \mathrm{E}[\mathbf{h}\mathbf{h}^H]$ and having the average power $\mathrm{tr}(\mathbf{R}_{\mathbf{hh}}) = N$:

$$M(\mathbf{Q}) = \mathrm{E}[(\hat{\mathbf{h}}^{mmse} - \mathbf{h})^H(\hat{\mathbf{h}}^{mmse} - \mathbf{h})] = \mathrm{E}[(\mathbf{Q}\mathbf{y}^p - \mathbf{h})^H(\mathbf{Q}\mathbf{y}^p - \mathbf{h})]. \tag{3.18}$$

It is shown in Appendix C.3 that minimisation of (3.18) with respect to $\mathbf{Q}$ yields

$$\hat{\mathbf{h}}^{\text{mmse}} = \sigma_{\text{p}}^{-2}\mathbf{R}_{\mathbf{hh}}\mathbf{C}^{\text{H}}(\mathbf{C}\mathbf{R}_{\mathbf{hh}}\mathbf{C}^{\text{H}} + SNR_{\text{p}}^{-1}\mathbf{I})^{-1}\mathbf{X}^{\mathbf{p}\,\text{H}}\mathbf{y}^{\mathbf{p}}, \tag{3.19}$$

where $SNR_{\text{p}} = \sigma_{\text{w}}^{-2}\sigma_{\text{p}}^{2}$.

It should be noted that, unlike the CLS estimator (3.8), which requires $P \geq L$, the MMSE solution (3.19) may exist even when $P < L$. However, the condition of $P \geq K$ should be satisfied to restore CFR without distortion in the absence of noise, where $K = \text{rk}(\mathbf{R}_{\mathbf{hh}}) \leq L$ is the number of multipath components, constituting the model order (see Subsections 2.2.1 and 2.2.2). This can easily be proved using the eigenvalue decomposition (EVD) of $\mathbf{C}\mathbf{R}_{\mathbf{hh}}\mathbf{C}^{\text{H}}$ and letting $SNR_{\text{p}} \to \infty$.

The MMSE estimator (3.19) relies on the statistical properties of the channel, namely on the noise variance σ_{w}^{2} and CFR correlation matrix $\mathbf{R}_{\mathbf{hh}}$. If the number of pilots is large, its implementation suffers from the considerable computational load inherent to the product with the $N \times P$ matrix $\mathbf{Q}$. One way to decrease complexity is to apply singular value decomposition (SVD) of $\mathbf{Q}$ as proposed by Edfors et al. [43]. Alternatively, we develop here transform-domain processing structure similar to the CLS implementation (Section 3.2).

In the derivation of the CLS estimator (3.8), we adopted the CIR finiteness assumption. As pointed out in Section 2.1, this deterministic property is a key requirement for the system design preventing interblock interference (IBI). Hence it is reasonable to imply this constraint on any channel estimation algorithm embedded in the receiver model. According to (3.2), $\mathbf{R}_{\mathbf{hh}}$ can be written in the following form:

$$\mathbf{R}_{\mathbf{hh}} = \text{E}[\mathbf{h}\mathbf{h}^{\text{H}}] = \text{E}[\mathbf{F}\mathbf{B}\underline{\mathbf{h}}(\mathbf{F}\mathbf{B}\underline{\mathbf{h}})^{\text{H}}] = \mathbf{F}\mathbf{B}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}, \tag{3.20}$$

where $\mathbf{R}_{\underline{\mathbf{hh}}} = \text{E}[\underline{\mathbf{h}}\underline{\mathbf{h}}^{\text{H}}]$ is the CIR correlation matrix (2.11).

Identity (3.20) leads to the DFT decomposition of the weighting matrix $\mathbf{Q}$ (refer to Appendix C.3 for details) that allows the CFR estimate to be expressed as

$$\hat{\mathbf{h}}^{\text{cmmse}} = \sigma_{\text{p}}^{-2}\mathbf{F}\mathbf{B}[\mathbf{I} - SNR_{\text{p}}^{-1}\mathbf{S}(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\text{p}}^{-1}\mathbf{S})^{-1}]\mathbf{S}\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}^{\text{H}}\mathbf{X}^{\mathbf{p}\,\text{H}}\mathbf{y}^{\mathbf{p}}, \tag{3.21}$$

where $\mathbf{S} = \left(\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}^{\text{H}}\mathbf{C}\mathbf{F}\mathbf{B}\right)^{-1}$, and the corresponding CIR estimate is

$$\hat{\underline{\mathbf{h}}}^{\text{cmmse}} = \sigma_{\text{p}}^{-2}[\mathbf{I} - SNR_{\text{p}}^{-1}\mathbf{S}(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\text{p}}^{-1}\mathbf{S})^{-1}]\mathbf{S}\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}^{\text{H}}\mathbf{X}^{\mathbf{p}\,\text{H}}\mathbf{y}^{\mathbf{p}}. \tag{3.22}$$

Due to the CIR finiteness assumption, we will call the estimator (3.21)-(3.22) *constrained MMSE* (CMMSE). Comparing (3.8) and (3.21), one can see that the CLS estimator represents a special case of CMMSE when $SNR_{\text{p}} \to \infty$. In fact, for any other SNR_{p} value the CMMSE estimator is obtained by *smoothing* the CLS CIR estimate through the product with the $L \times L$ matrix $\mathbf{I} - SNR_{\text{p}}^{-1}\mathbf{S}(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\text{p}}^{-1}\mathbf{S})^{-1}$.

When the pilot subcarriers are equispaced, formula (3.21) is simplified to $\hat{\underline{\mathbf{h}}}^{\text{cmmse}} = \sigma_{\text{p}}^{-2}P^{-1}[\mathbf{I} - P^{-1}SNR_{\text{p}}^{-1}(\mathbf{R}_{\underline{\mathbf{hh}}} + P^{-1}SNR_{\text{p}}^{-1}\mathbf{I})^{-1}]\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}^{\text{H}}\mathbf{X}^{\mathbf{p}\,\text{H}}\mathbf{y}^{\mathbf{p}}$.

The CMMSE estimator (3.21) uses a priori information about the noise variance σ_{w}^2 and CIR correlation matrix $\mathbf{R}_{\underline{\mathbf{hh}}}$, and is optimal when these statistics are perfectly known at the receiver. In practice, $\mathbf{R}_{\underline{\mathbf{hh}}}$ and σ_{w}^2 need to be acquired themselves. In general, this can be accomplished using external methods.

The correlation matrix of the channel response can be estimated by the base station equipment, which exploits higher-complexity processing algorithms in comparison with the subscriber radios, and signalled to the user receiver. The applicability of this method is grounded by the channel reciprocity property [16] when the downlink and uplink propagation links, operating in the time division duplex mode, have the same statistical characteristics.

The noise power σ_{w}^2 can be a priori known in a number of cases, e.g., determined as a function of the amplifier gain at the receiver input. If, however, knowledge of σ_{w}^2 is not available, it can be replaced by its estimate $\hat{\sigma}_{\mathrm{w}}^2$. The simplest (and the most reliable) method to obtain that estimate is straightforward σ_{w}^2 measurement on one or several non-active (virtual) subcarriers [46]. The *virtual subcarriers* (VCs) are often present in the standardised OFDM systems, being located at DC and spectrum edges [20].

3.3.2 Recursive operation mode

Even when the true values of the design parameters, $\mathbf{R}_{\underline{\mathbf{hh}}}$ and σ_{w}^2, are available and regularly updated at the receiver, the CMMSE estimator (3.21) requires real-time inversion of the correlation matrix $\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\mathrm{p}}^{-1}\mathbf{S}$ for optimal operation that is found to be impractical due to considerable increase of the receiver's complexity. In this subsection, we develop methods of a computationally efficient design of the CMMSE estimator without matrix inversions. The proposed recursive algorithm performs internal estimation and real-time block-by-block updates of the second-order channel statistics.

3.3.2.1 Reduced-complexity recursive CMMSE algorithm

In Subsection 3.3.1, it has been established that the CMMSE estimator can be viewed as a cascade of the CLS estimator and the smoothing filter, i.e.

$$\hat{\underline{\mathbf{h}}}_m^{\mathrm{cmmse}} = \mathbf{W}\tilde{\underline{\mathbf{h}}}_m, \tag{3.23}$$

where $\mathbf{W} = \mathbf{I} - SNR_{\mathrm{p}}^{-1}\mathbf{S}(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\mathrm{p}}^{-1}\mathbf{S})^{-1}$.

Hence the correlation matrix of the smoothing filter input is given by:

$$\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}} = \mathrm{E}[\tilde{\underline{\mathbf{h}}}_m \tilde{\underline{\mathbf{h}}}_m^{\mathrm{H}}] = \mathrm{E}[\underline{\mathbf{h}}_m \underline{\mathbf{h}}_m^{\mathrm{H}}] + \mathrm{E}[\tilde{\underline{\boldsymbol{\varepsilon}}}_m \tilde{\underline{\boldsymbol{\varepsilon}}}_m^{\mathrm{H}}] = \mathbf{R}_{\underline{\mathbf{hh}}} + \mathbf{R}_{\underline{\tilde{\boldsymbol{\varepsilon}}}\underline{\tilde{\boldsymbol{\varepsilon}}}}, \tag{3.24}$$

where $\tilde{\underline{\boldsymbol{\varepsilon}}}_m = \sigma_{\mathrm{p}}^{-2}\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\,\mathrm{H}}\mathbf{w}_m^{\mathbf{p}}$ (3.11), $\mathrm{E}[\tilde{\underline{\boldsymbol{\varepsilon}}}_m \underline{\mathbf{h}}_m^{\mathrm{H}}] = \mathrm{E}[\mathbf{h}_m \tilde{\underline{\boldsymbol{\varepsilon}}}_m^{\mathrm{H}}] = \mathbf{0}_{L\times L}$, and $\mathbf{R}_{\underline{\tilde{\boldsymbol{\varepsilon}}}\underline{\tilde{\boldsymbol{\varepsilon}}}} = SNR_{\mathrm{p}}^{-1}\mathbf{S}$ according to (3.12).

The key idea of the reduced-complexity design is recursive estimation of the precision matrix $\mathbf{T}_{\underline{\tilde{h}}\tilde{h}} = \mathbf{R}_{\underline{\tilde{h}}\tilde{h}}^{-1}$ instead of the direct inversion of the $\mathbf{R}_{\underline{\tilde{h}}\tilde{h}}$ estimate. Note that existence of the inverse of the positive definite Hermitean matrix $\mathbf{R}_{\underline{\tilde{h}}\tilde{h}}$ is always guaranteed as its eigenvalues are all positive for $\sigma_{\mathrm{w}}^2 > 0$.

Replacing $\mathbf{R}_{\underline{\tilde{h}}\tilde{h}}$ with its estimate, computed for the mth block using the *exponential averaging* technique, yields:

$$\hat{\mathbf{R}}_{\underline{\tilde{h}}\tilde{h}m} = (1-\alpha_{\mathrm{FF}})\hat{\mathbf{R}}_{\underline{\tilde{h}}\tilde{h}m-1} + \alpha_{\mathrm{FF}}\underline{\tilde{\mathbf{h}}}_m\underline{\tilde{\mathbf{h}}}_m^{\mathrm{H}}, \tag{3.25}$$

where α_{FF} is the *forgetting factor* (FF), satisfying the condition $0 < \alpha_{\mathrm{FF}} < 1$.

The estimate of the precision matrix $\mathbf{T}_{\underline{\tilde{h}}\tilde{h}}$ is found applying the matrix inversion lemma [102] to (3.25):

$$\hat{\mathbf{T}}_{\underline{\tilde{h}}\tilde{h}m} = \left[(1-\alpha_{\mathrm{FF}})\hat{\mathbf{R}}_{\underline{\tilde{h}}\tilde{h}m-1} + \alpha_{\mathrm{FF}}\underline{\tilde{\mathbf{h}}}_m\underline{\tilde{\mathbf{h}}}_m^{\mathrm{H}}\right]^{-1} = \frac{1}{(1-\alpha_{\mathrm{FF}})}\hat{\mathbf{T}}_{\underline{\tilde{h}}\tilde{h}m-1}\left[\mathbf{I} - \frac{\underline{\tilde{\mathbf{h}}}_m\underline{\tilde{\mathbf{h}}}_m^{\mathrm{H}}\hat{\mathbf{T}}_{\underline{\tilde{h}}\tilde{h}m-1}}{(1-\alpha_{\mathrm{FF}})/\alpha_{\mathrm{FF}} + \mathrm{tr}(\underline{\tilde{\mathbf{h}}}_m\underline{\tilde{\mathbf{h}}}_m^{\mathrm{H}}\hat{\mathbf{T}}_{\underline{\tilde{h}}\tilde{h}m-1})}\right], \tag{3.26}$$

where the initial setting of the precision matrix, corresponding to $m = -1$, should be chosen as $\hat{\mathbf{T}}_{\underline{\tilde{h}}\tilde{h}-1} = \iota\mathbf{I}$, where $\iota \to \infty$ (in practice, equal to some big enough value) minimises average's bias towards the initial setting. It can also be interpreted as the setting, which maximises rapidity of $\hat{\mathbf{T}}_{\underline{\tilde{h}}\tilde{h}}$ change at the first update, expressed by $\lim_{\iota\to\infty}\mathrm{tr}[(\iota\mathbf{I} - \hat{\mathbf{T}}_{\underline{\tilde{h}}\tilde{h}0})^{\mathrm{H}}(\iota\mathbf{I} - \hat{\mathbf{T}}_{\underline{\tilde{h}}\tilde{h}0})] \to \infty$.

Hence the weighting matrix $\mathbf{W}$ in (3.23) is described recursively as $\mathbf{W}_m = \mathbf{I} - SNR_{\mathrm{p}}^{-1}\mathbf{S}\hat{\mathbf{T}}_{\underline{\tilde{h}}\tilde{h}m}$, leading to the *recursive CMMSE* (RMMSE) estimator

$$\hat{\mathbf{h}}_m^{\mathrm{rmmse}} = \mathbf{F}\mathbf{B}\hat{\underline{\mathbf{h}}}_m^{\mathrm{rmmse}} = \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}(\mathbf{I} - SNR_{\mathrm{p}}^{-1}\mathbf{S}\hat{\mathbf{T}}_{\underline{\tilde{h}}\tilde{h}m})\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\,\mathrm{H}}\mathbf{y}_m^{\mathbf{p}}, \tag{3.27}$$

whose functional diagram is shown in Fig.3.2.

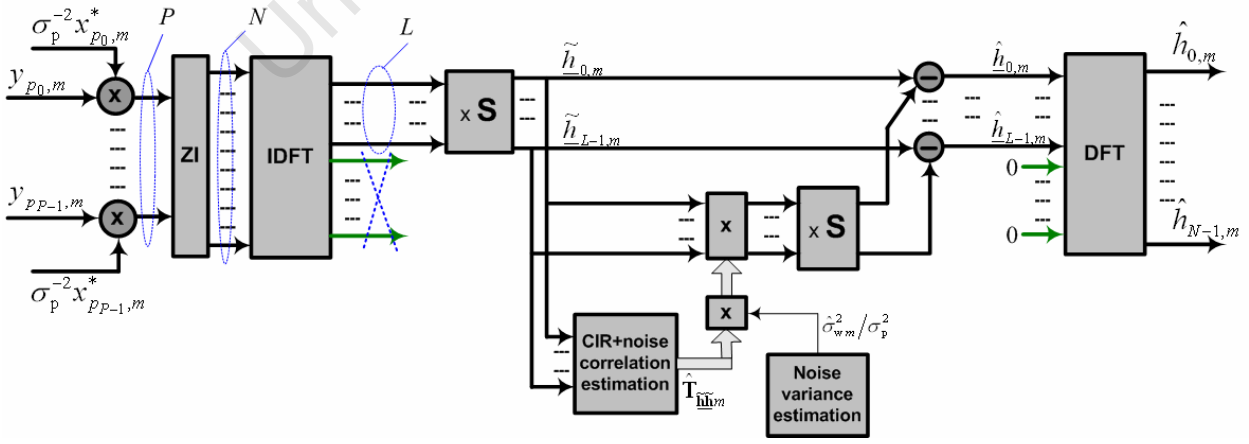


Fig.3.2. RMMSE estimator

An intent look on the principle of the derived RMMSE estimator reveals similarity with the recursive least squares (RLS) algorithm [135]. However, our derivation is associated with the MMSE filter formulation, not LS. In particular, it is evident from (3.27) that in the steady state the RMMSE solution converges to CMMSE due to $\lim_{m\to\infty}(\hat{\mathbf{T}}_{\underline{\tilde{h}}\underline{\tilde{h}}\,m}) \to \mathbf{T}_{\underline{\tilde{h}}\underline{\tilde{h}}}$.

Note that the algorithm (3.26)-(3.27) does not involve any matrix inversions. The matrix $\mathbf{S}$ is fixed by the design associated with a specific pilot arrangement. Estimation of the precision matrix $\hat{\mathbf{T}}_{\underline{\tilde{h}}\underline{\tilde{h}}\,m}$ is of reasonable complexity due to the small ($L\times L$) size of $\hat{\mathbf{T}}_{\underline{\tilde{h}}\underline{\tilde{h}}\,m}$ (typically $L \le N/4$). If pilot subcarriers in the OFDM spectrum are equispaced, then $\mathbf{S} = P^{-1}\mathbf{I}$ and hence the CIR estimate in (3.27) is considerably simplified to $\hat{\underline{\mathbf{h}}}_m^{\text{rmmse}} = \sigma_{\text{p}}^{-2}P^{-1}[\mathbf{I} - (P\,SNR_{\text{p}})^{-1}\hat{\mathbf{T}}_{\underline{\tilde{h}}\underline{\tilde{h}}\,m}]\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}^{\text{H}}\mathbf{X}^{\text{p H}}\mathbf{y}_m^{\text{p}}$.

It can be seen from (3.27) that the only unknown design parameter for the RMMSE estimator is the noise variance σ_{w}^2. It has already been mentioned that in the absence of the noise power information, σ_{w}^2 can be measured on VCs, yielding the estimate $\hat{\sigma}_{\text{w}\,m}^2$. Despite the VC-inherent loss in spectral efficiency, an advantage of this method is decoupling of the noise variance estimation from the channel response estimation. It ensures better stability of the RMMSE algorithm, especially in case of the non-stationary operational conditions characterised by σ_{w}^2 fluctuations. Due to statistical independence of the WGN samples, affecting successive OFDM blocks, one can use averaging with relatively short memory in order to compute $\hat{\sigma}_{\text{w}\,m}^2$ based on per-block VC power measurements.

Finally, note that although the RMMSE estimator uses a sequence of blocks to compute CIR-plus-noise precision matrix (3.26), it still produces intrablock estimates of the channel response as the nature of the interblock channel variation is not taken into account.

3.3.2.2 Switching between estimator modes

System initialisation stage, as well as non-stationarities in channel (e.g., sudden changes in the CIR correlation matrix) can potentially cause instability of the RMMSE algorithm, which is manifested as a dramatic MSE increase. In such situation, it is desirable to perform immediate switching to the CLS mode by means of zeroing σ_{w}^2 (i.e. SNR_{p}^{-1}) in (3.27).

We propose to use cross-correlation between the input of the smoothing module of the RMMSE algorithm, $\tilde{\underline{\mathbf{h}}}_m$, and the input-output difference, $\underline{\xi}_m = \tilde{\underline{\mathbf{h}}}_m - \hat{\underline{\mathbf{h}}}_m^{\text{rmmse}} = SNR_{\text{p}}^{-1}\mathbf{S}\,\hat{\mathbf{T}}_{\underline{\tilde{h}}\underline{\tilde{h}}\,m}\tilde{\underline{\mathbf{h}}}_m$, as a criterion for switching estimator's modes:

$$\mathbf{R}_{\underline{\xi}\underline{\tilde{h}}\,m} = \text{E}[\underline{\xi}_m\tilde{\underline{\mathbf{h}}}_m^{\text{H}}] = SNR_{\text{p}}^{-1}\mathbf{S}\,\hat{\mathbf{T}}_{\underline{\tilde{h}}\underline{\tilde{h}}\,m}\mathbf{R}_{\underline{\tilde{h}}\underline{\tilde{h}}}. \tag{3.28}$$

As the Hermitean matrices $\mathbf{S}$, $\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}$ and $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}$ are all positive-definite, $\mathrm{tr}(\mathbf{R}_{\underline{\xi}\underline{\tilde{\mathbf{h}}}m}) > 0$. In the optimal case of $\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m} = \mathbf{T}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}$, $\mathrm{tr}(\mathbf{R}_{\underline{\xi}\underline{\tilde{\mathbf{h}}}m}) = SNR_{\mathrm{p}}^{-1}\mathrm{tr}(\mathbf{S})$. If, however, $\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m} \neq \mathbf{T}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}$, a dramatic increase of $\mathrm{tr}(\mathbf{R}_{\underline{\xi}\underline{\tilde{\mathbf{h}}}m})$ over the nominal value might be expected.

Replacing expectation in (3.28) with the exponential averaging, producing an estimate $\hat{\mathbf{R}}_{\underline{\xi}\underline{\tilde{\mathbf{h}}}m}$ on the analogy of (3.25), yields a practical expression, which can be used to switch between the robust (CLS) and the recursive modes:

$$\mathrm{tr}(\hat{\mathbf{R}}_{\underline{\xi}\underline{\tilde{\mathbf{h}}}m}) \gtrless \zeta\, SNR_{\mathrm{p}}^{-1}\mathrm{tr}(\mathbf{S}), \tag{3.29}$$

where the tolerance factor ζ is needed to account for $\mathrm{tr}(\hat{\mathbf{R}}_{\underline{\xi}\underline{\tilde{\mathbf{h}}}m})$ variance due to the short averaging memory, and σ_{w}^2 is either known a priori or measured on VCs as $\hat{\sigma}_{\mathrm{w}m}^2$. The estimator operates recursively, while the sign "<" holds in condition (3.29). When it changes to "≥", CLS mode has to be enabled.

In some cases (e.g., repeated initialisations of the estimator) introduction of distinct tolerance factors for switching from the robust to the recursive mode (ζ_1) and back (ζ_2) can improve triggering accuracy. $\zeta_2 > \zeta_1$ setting ensures transition to the recursive operation once $\mathbf{T}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}$ and σ_{w}^2 have been estimated with the required accuracy level and tolerance to $\mathrm{tr}(\hat{\mathbf{R}}_{\underline{\xi}\underline{\tilde{\mathbf{h}}}m})$ surges while in the recursive mode.

3.3.2.3 Forgetting factor adaptation

Forgetting factor α_{FF} in (3.26) is assumed to be preset to some known value, ensuring efficient averaging with tolerable fluctuation of $\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}$. For the fixed α_{FF}, $\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}$ variance depends on the rapidity of the CIR change between the successive blocks, i.e. the maximum Doppler shift ω_{D} (2.28). Thus, to ensure accurate $\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}$ estimation, one has to select α_{FF} in accordance with ω_{D}. Higher ω_{D} values ($\omega_{\mathrm{D}} \to \pi$) allow for larger α_{FF} settings, whereas smaller ω_{D} values ($\omega_{\mathrm{D}} \to 0$), characterising strong channel response correlation in the time domain, require averaging over a wider range of observations that is possible when α_{FF} is small enough. Derivation of the optimal FF setting for the LMS and RLS estimators, depending on the time variation statistics of the channel, can be found in the works [88] and [99]. In most cases, however, accurate information about ω_{D} is not available. Therefore the design of the algorithm (3.26)-(3.27) is challenged by the question: what should the FF be equal to? As shown by Song et al. [136], introduction of the variable forgetting factor, adapting itself to variation of the estimated process, could appear a solution to the problem. Here we propose such an adaptation mechanism. The developed algorithm is related to the normalised least mean squares (NLMS) principle. In general, our idea is close to [136], but the algorithm is entirely different.

The exponential averaging procedure, yielding an estimate $\hat{g}_m$ on some interval $(m-I+1)$ to m, can be interpreted as passing the input samples $\hat{g}_{m-I}$, u_{m-I+1}, …, u_{m-1}, u_m through the FIR structure with the tap-weights equal to $(1-\alpha_{\mathrm{FF}})^I$, $\alpha_{\mathrm{FF}}(1-\alpha_{\mathrm{FF}})^{I-1}$,..., $\alpha_{\mathrm{FF}}(1-\alpha_{\mathrm{FF}})$, and α_{FF} correspondingly, as described by

$$\hat{g}_m = (1-\alpha_{\mathrm{FF}})\,\hat{g}_{m-1} + \alpha_{\mathrm{FF}} u_m = (1-\alpha_{\mathrm{FF}})^I\,\hat{g}_{m-I} + \alpha_{\mathrm{FF}} \sum_{k=0}^{I-1} (1-\alpha_{\mathrm{FF}})^k u_{m-k} \ . \tag{3.30}$$

Tab.3.1. Forgetting factor adaptation algorithm

Initial conditions:

Set $\alpha_{\mathrm{FF}}(-1)$ to satisfy $0 < \alpha_{\mathrm{FF}}(-1) < 1$

Set $\hat{g}_{l,-1} := 1/\iota$, $l \in [0, L-1]$

Set I (processing frame length)

$m := 0$

Recursion body:

If $(m+1)$ mod $I = 0$ do

(a) $$f_l := \mu\, u_{l,m} \Big/ \left[\hat{g}^2_{l,m-I} + \sum_{k=0}^{I-1} u^2_{l,m-k} \right],$$

where $0 < \mu < 2$ (typically $\mu \to 2$ is selected)

(b) $$\delta_l := \left| \hat{g}_{l,m} - \hat{g}_{l,m-I} \right| \Big/ \hat{g}_{l,m}$$

(c) $$\Delta_l := \begin{cases} -f_l \delta_l \hat{g}_{l,m}, & \text{if } \delta_l > \upsilon \text{ and } \alpha_{\mathrm{FF}}(m-1) > f_l \delta_l \hat{g}_{l,m} \\ 0, & \text{if } \delta_l > \upsilon \text{ and } \alpha_{\mathrm{FF}}(m-1) \le f_l \delta_l \hat{g}_{l,m} \\ f_l \nu \hat{g}_{l,m}, & \text{if } \delta_l \le \upsilon \end{cases},$$

where $0 < \upsilon < 1$ ($\upsilon << 1$), $0 < \nu < 1$ ($\nu << \upsilon$)

(d) $$\alpha_{\mathrm{FF}}(m) := \alpha_{\mathrm{FF}}(m-1) + \frac{1}{L} \sum_{l=0}^{L-1} \Delta_l$$

else $\alpha_{\mathrm{FF}}(m) := \alpha_{\mathrm{FF}}(m-1)$

Return

$m := m+1$

Go to beginning of the recursion body

The basic idea of the proposed adaptation technique is to adjust the tap-weight α_{FF} (other tap-weights are dependent on it) until the relative error between $\hat{g}_m$ and $\hat{g}_{m-I}$ falls under the specified precision threshold υ. Initially α_{FF} is set to some large value lesser than unity. During the adaptation, α_{FF} is decremented by the term proportional to $\left| \hat{g}_m - \hat{g}_{m-I} \right|$. After the precision threshold is met, the algorithm tries to increment α_{FF} by a value,

which is an order of magnitude smaller than υ. Thus, for the given update interval *I*, the adaptation algorithm finds the largest α_{FF} (corresponding to the minimal averaging memory), satisfying the selected precision criterion. The overall algorithm, applied to $\hat{\tilde{\mathbf{p}}}_m = \mathrm{dg}(\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\tilde{\mathbf{h}}m})$ estimation, is listed in Tab.3.1, where we let $\hat{g}_{l,m} = [\mathrm{dg}(\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\tilde{\mathbf{h}}m})]_{l,l}$, $u_{l,m} = [\underline{\tilde{\mathbf{h}}}_m \underline{\tilde{\mathbf{h}}}_m^{\mathrm{H}}]_{l,l}$, $l \in [0, L-1]$, and compute $\alpha_{\mathrm{FF}}(m)$ as a mean of FF estimates obtained individually for each $\hat{g}_{l,m}$.

With regard to the design complexity, it should be noted that α_{FF} adaptation can be performed on the base station side, and then the estimated α_{FF} can be signalled back to the subscriber station. If the uplink and the downlink channels are operating at different RF frequencies, then α_{FF} can be predicted for the reverse channel, taking into account ω_{D} proportionality to the RF frequency in (2.28). From this standpoint it has relatively no significance what kind of FF adaptation mechanism is used. We do not claim optimality of the proposed method, and this topic is outside of the scope of the thesis.

3.3.3 Computational complexity

The transform-domain architecture of the CMMSE and RMMSE estimators requires a different number of mathematical operations in comparison with CLS only in the filtering part. The pre- and post-filtering stages are still characterised by $C_{\mathrm{M}}^{\mathrm{FFT}}$ multiplies and $C_{\mathrm{A}}^{\mathrm{FFT}}$ additions as described in Subsection 3.2.2.

If the CMMSE estimator is characterised by the preset $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\tilde{\mathbf{h}}} = \mathbf{R}_{\underline{\mathbf{h}}\mathbf{h}} + SNR_{\mathrm{p}}^{-1}\mathbf{S}$ setting in (3.21), i.e. the CIR-plus-noise correlation matrix $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\tilde{\mathbf{h}}}$ is not updated in the operation process, the multiplicative and additive complexities are equivalent to the CLS case and are given by (3.9) and (3.10). If the estimate of $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\tilde{\mathbf{h}}}$ is acquired and block-wise updated by external means, local inversion is needed to compute $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\tilde{\mathbf{h}}}^{-1}$ for the given SNR regime. As mentioned in Appendix B.1, the efficient and reliable inversion procedure is Gauss-Jordan elimination, which requires L^3 multiplies and $L^3 - 2L^2 + L$ additions. Some, but insignificant reduction of the computational effort is possible when using QRD instead, in analogy with the method in Appendix B.3. However, QRD is less numerically stable. The remaining mathematical operations represent product of the L-size vector with $\mathbf{S}$, which is analogous to the CLS algorithm, and the subsequent products with $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\tilde{\mathbf{h}}}^{-1}$ and $SNR_{\mathrm{p}}^{-1}\mathbf{S}$, involving L^2 multiplies and $L(L-1)$ additions each. Apart from that, the scalar-matrix product $SNR_{\mathrm{p}}^{-1}\mathbf{S}$ (where $\mathbf{S}$ is doubly symmetric) and the sum $\mathbf{R}_{\underline{\mathbf{h}}\mathbf{h}} + SNR_{\mathrm{p}}^{-1}\mathbf{S}$ require $\sum_{k=0}^{L/2}(L-2k) = \frac{L}{2}\left(\frac{L}{2}+1\right)$ multiplications and $\sum_{k=1}^{L} k = \frac{L}{2}(L+1)$ additions, respectively (assuming even L). Finally, the difference $\mathbf{I} - SNR_{\mathrm{p}}^{-1}\mathbf{S}\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\tilde{\mathbf{h}}}^{-1}$ is viewed as L additions.

For the RMMSE estimator, computation of $\mathbf{W}_m = \mathbf{I} - \dfrac{\underline{\tilde{\mathbf{h}}}_m \underline{\tilde{\mathbf{h}}}_m^{\mathrm{H}} \hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m-1}}{(1-\alpha_{\mathrm{FF}})/\alpha_{\mathrm{FF}} + \mathrm{tr}(\underline{\tilde{\mathbf{h}}}_m \underline{\tilde{\mathbf{h}}}_m^{\mathrm{H}} \hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m-1})}$ in (3.26) involves $2L^2 + L$ multiplies and $L(L-1) + 2L$ additions. The smoothing module of (3.27), which represents a product with the matrix $SNR_{\mathrm{p}}^{-1}(1-\alpha_{\mathrm{FF}})^{-1}\mathbf{S}\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m-1}\mathbf{W}_m\mathbf{S}$, has the multiplicative complexity of $4L^2 + 2L$ and the additive complexity of $4L(L-1)$ operations.

The overall complexity for different CMMSE estimator variants is shown in Tab.3.2. Here the mechanisms of SNR_{p}^{-1} and α_{FF} estimation are not taken into account.

One can see an order of a magnitude reduction in both the multiplicative and additive complexity of the RMMSE in comparison with the CMMSE driven by the estimated CIR correlation matrix, due to avoidance of the matrix inversion. It should also be noted that the equispaced pilot pattern remarkably lowers computational burden of the fixed CMMSE and RMMSE estimators.

Tab.3.2. Computational complexity for the intrablock CMMSE estimator family

Estimator	**Multiplications** ($C_{\mathrm{M}}^{\mathrm{cmmse}}$)	**Additions** ($C_{\mathrm{A}}^{\mathrm{cmmse}}$)
CMMSE (fixed $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}$)	$C_{\mathrm{M}}^{\mathrm{FFT}} + L^2$	$C_{\mathrm{A}}^{\mathrm{FFT}} + L^2 - L$
CMMSE (fixed $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}$) with comb pilots	$C_{\mathrm{M}}^{\mathrm{FFT}}$	$C_{\mathrm{A}}^{\mathrm{FFT}}$
CMMSE (updated $\mathbf{R}_{\underline{\mathbf{h}}\underline{\mathbf{h}}}$)	$C_{\mathrm{M}}^{\mathrm{FFT}} + L^3 + \frac{13}{4}L^2 + \frac{1}{2}L$	$C_{\mathrm{A}}^{\mathrm{FFT}} + L^3 + \frac{3}{2}L^2 - \frac{1}{2}L$
CMMSE (updated $\mathbf{R}_{\underline{\mathbf{h}}\underline{\mathbf{h}}}$) with comb pilots	$C_{\mathrm{M}}^{\mathrm{FFT}} + L^3 + L^2 + L$	$C_{\mathrm{A}}^{\mathrm{FFT}} + L^3 - L^2 + 2L$
RMMSE	$C_{\mathrm{M}}^{\mathrm{FFT}} + 6L^2 + 3L$	$C_{\mathrm{A}}^{\mathrm{FFT}} + 5L^2 - 2L$
RMMSE with comb pilots	$C_{\mathrm{M}}^{\mathrm{FFT}} + 4L^2 + 3L$	$C_{\mathrm{A}}^{\mathrm{FFT}} + 3L^2$

3.3.4 MSE analysis

In this subsection, we will derive MSE of the CMMSE channel estimator with the optimal and suboptimal design when the statistical parameters, describing the channel, are known only partially. The same MSE performance results correspond to the RMMSE algorithm if it is analysed in the steady state, due to $\lim_{m\to\infty}(\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m}) \to \mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1}$ (refer to Subsection 3.3.2).

Let $\mathbf{R}_{\underline{\mathbf{h}}\underline{\mathbf{h}}}$ and σ_{w}^2 be the true CIR correlation matrix and noise variance, and $\overline{\mathbf{R}}_{\underline{\mathbf{h}}\underline{\mathbf{h}}}$ and $\overline{\sigma}_{\mathrm{w}}^2$ denote CIR correlation matrix and noise variance used for the design of the estimator (3.21), namely

$$\hat{\mathbf{h}} = \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}(\mathbf{I} - \overline{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1})\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\mathrm{H}}\mathbf{y}^{\mathbf{p}}, \tag{3.31}$$

where $\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}} = \overline{\mathbf{R}}_{\underline{\mathbf{h}}\underline{\mathbf{h}}} + \overline{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}$.

Correlation matrix of the CFR estimation error, $\boldsymbol{\varepsilon} = \hat{\mathbf{h}} - \mathbf{h} = \mathbf{F}\mathbf{B}(\hat{\underline{\mathbf{h}}} - \underline{\mathbf{h}}) = \mathbf{F}\mathbf{B}\underline{\boldsymbol{\varepsilon}}$, can be expressed as

$$\mathbf{R}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} = \mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{H}}] = \mathbf{F}\mathbf{B}\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}, \tag{3.32}$$

where $\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}} = \mathrm{E}[\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}^{\mathrm{H}}] = \mathrm{E}[(\hat{\underline{\mathbf{h}}} - \underline{\mathbf{h}})(\hat{\underline{\mathbf{h}}} - \underline{\mathbf{h}})^{\mathrm{H}}]$ is the correlation matrix of the CIR estimation error. Derivation of $\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}$ is presented in Appendix C.4, resulting in

$$\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}} = \mathbf{R}_{\underline{\tilde{\boldsymbol{\varepsilon}}}\underline{\tilde{\boldsymbol{\varepsilon}}}} - \mathbf{R}_{\underline{\boldsymbol{\eta}}\underline{\boldsymbol{\eta}}}, \tag{3.33}$$

where $\mathbf{R}_{\underline{\tilde{\boldsymbol{\varepsilon}}}\underline{\tilde{\boldsymbol{\varepsilon}}}} = SNR_{\mathrm{p}}^{-1}\mathbf{S}$ is the correlation matrix of the CIR estimation error, corresponding to the CLS estimator, $\mathbf{R}_{\underline{\boldsymbol{\eta}}\underline{\boldsymbol{\eta}}} = S\overline{NR}_{\mathrm{p}}^{-1}\mathbf{S}\,\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1/2}(2\,SNR_{\mathrm{p}}^{-1}\mathbf{I} - S\overline{NR}_{\mathrm{p}}^{-1}\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1/2}\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1/2})\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1/2}\mathbf{S}$ is the covariance reduction term for $\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}$, $S\overline{NR}_{\mathrm{p}} = \overline{\sigma}_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^{2}$, and $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}} = \mathbf{R}_{\underline{\mathbf{h}}\underline{\mathbf{h}}} + \mathbf{R}_{\underline{\tilde{\boldsymbol{\varepsilon}}}\underline{\tilde{\boldsymbol{\varepsilon}}}}$.

MSE, averaged over all the subcarriers in the band, is subsequently computed as

$$\sigma_{\varepsilon}^{2} = N^{-1}\mathrm{tr}(\mathbf{R}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}) = \mathrm{tr}(\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}) = \mathrm{tr}(\mathbf{R}_{\underline{\tilde{\boldsymbol{\varepsilon}}}\underline{\tilde{\boldsymbol{\varepsilon}}}}) - \mathrm{tr}(\mathbf{R}_{\underline{\boldsymbol{\eta}}\underline{\boldsymbol{\eta}}}) = \sigma_{\tilde{\varepsilon}}^{2} - \sigma_{\eta}^{2}. \tag{3.34}$$

The positive semi-definiteness of the correlation matrix $\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}$ (mandatory property) is ensured if and only if $\mathrm{tr}(\mathbf{R}_{\underline{\boldsymbol{\eta}}\underline{\boldsymbol{\eta}}}) \le \mathrm{tr}(\mathbf{R}_{\underline{\tilde{\boldsymbol{\varepsilon}}}\underline{\tilde{\boldsymbol{\varepsilon}}}})$, where $\mathbf{R}_{\underline{\tilde{\boldsymbol{\varepsilon}}}\underline{\tilde{\boldsymbol{\varepsilon}}}}$ is always positive definite and $\mathbf{R}_{\underline{\boldsymbol{\eta}}\underline{\boldsymbol{\eta}}}$ has to be positive semi-definite for the (sub)optimal estimator design. The latter statement follows from the condition that the correlation matrix has non-negative eigenvalues, and guarantees performance to be no worse than that of the CLS estimator.

Note that CFR estimation error variances, being the diagonal elements of $\mathbf{R}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}$ (3.32), are identical for all the subcarriers only when $\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}$ is equal to a scaled identity matrix, i.e. $\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}} = \beta\mathbf{I}$. In such a case MSE at the nth subcarrier is expressed as $\sigma_{\varepsilon n}^{2} = \beta\left\|((\mathbf{F}\mathbf{B})^{\mathrm{T}})_{n}\right\|^{2} = \beta L$ for $\forall n \in [0, N-1]$. In further examples, it will be pointed out when this condition holds (e.g., for the CLS operation mode with equispaced pilots, or for the optimally designed CMMSE with equispaced pilots when the sample-spaced channel correlation matrix has equal eigenvalues). This circumstance should be taken into account in the adaptively loaded systems [15] as the probability of detection error on a selected subcarrier is influenced by the channel estimation error observed on this subcarrier.

In Appendix C.5, we show that the minimum MSE (3.34), at least for the CLS operational mode, is attained if the pilot subcarriers are equispaced. In this connection the following inequality holds:

$$\sigma_{\tilde{\varepsilon}}^{2} \ge SNR_{\mathrm{p}}^{-1}P^{-1}L \tag{3.35}$$

that turns into the equality in case of the comb pilot pattern.

From (3.31), in analogy with the derivation in Appendix C.4, one can express the cross-correlation of CFR and estimation error as

$$\begin{aligned}\sigma_{\varepsilon\mathrm{h}}^{2} &= N^{-1}\mathrm{E}[\mathbf{h}^{\mathrm{H}}\boldsymbol{\varepsilon}] \\ &= \mathrm{tr}[\mathrm{E}(\hat{\underline{\mathbf{h}}}\underline{\mathbf{h}}^{\mathrm{H}}) - \mathrm{E}(\underline{\mathbf{h}}\underline{\mathbf{h}}^{\mathrm{H}})] \\ &= \mathrm{tr}[(\mathbf{I} - \overline{\sigma}_{\mathrm{w}}^{2}\sigma_{\mathrm{p}}^{-2}\mathbf{S}\,\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1})\mathbf{R}_{\underline{\mathbf{h}}\underline{\mathbf{h}}} - \mathbf{R}_{\underline{\mathbf{h}}\underline{\mathbf{h}}}] \\ &= -\overline{\sigma}_{\mathrm{w}}^{2}\sigma_{\mathrm{p}}^{-2}\mathrm{tr}[\mathbf{S}(\overline{\mathbf{R}}_{\underline{\mathbf{h}}\underline{\mathbf{h}}} + \overline{\sigma}_{\mathrm{w}}^{2}\sigma_{\mathrm{p}}^{-2}\mathbf{S})^{-1}\mathbf{R}_{\underline{\mathbf{h}}\underline{\mathbf{h}}}].\end{aligned} \tag{3.36}$$

It is obvious that $\sigma^2_{\tilde{\varepsilon}\tilde{h}} = 0$ if and only if $\bar{\sigma}^2_{w} = 0$, i.e. when the estimator is designed as CLS. This conclusion is important if one considers a cascade of two linear estimators. For the cascaded structure to be optimal, an error at the output of the primary estimator must be uncorrelated with the variable of interest (true channel response). As one can see, this condition is satisfied if the primary estimator is designed as CLS.

We now consider several practical approaches towards the optimal and suboptimal design of the CMMSE estimator, based on $\bar{\mathbf{R}}_{\underline{\mathbf{hh}}}$ and $\bar{\sigma}^2_{w}$ selection.

3.3.4.1 CLS mode

In this special case, the estimator (3.31) is transformed into the CLS estimator by setting $\bar{\sigma}^2_{w} = 0$:

$$\hat{\mathbf{h}}^{\text{cls}} = \mathbf{F}\mathbf{B}\hat{\underline{\mathbf{h}}}^{\text{cls}} = \sigma_p^{-2}\mathbf{F}\mathbf{B}\mathbf{S}\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}^{\text{H}}\mathbf{X}^{\mathbf{p}\,\text{H}}\mathbf{y}^{\mathbf{p}}. \tag{3.37}$$

It has already been established in Subsection 3.2.3 that the corresponding correlation matrix of the CFR estimation error, $\tilde{\boldsymbol{\varepsilon}} = \hat{\mathbf{h}}^{\text{cls}} - \mathbf{h}$, and MSE are $\mathbf{R}_{\tilde{\varepsilon}\tilde{\varepsilon}} = \mathbf{F}\mathbf{B}\mathbf{R}_{\underline{\tilde{\varepsilon}\tilde{\varepsilon}}}\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}$ (3.12) and $\sigma^2_{\tilde{\varepsilon}} = (s\,SNR_p)^{-1}$ (3.13), where $\mathbf{R}_{\underline{\tilde{\varepsilon}\tilde{\varepsilon}}} = SNR_p^{-1}\mathbf{S}$ and $s = 1/\text{tr}(\mathbf{S})$. For the equispaced arrangement of the pilot subcarriers, the MSE expression is reduced to $\sigma^2_{\tilde{\varepsilon},\text{es}} = SNR_p^{-1}P^{-1}L$ (3.15).

Note that $\mathbf{R}_{\underline{\tilde{\varepsilon}\tilde{\varepsilon}}}$ becomes diagonal for the equispaced pilot pattern: $\mathbf{R}_{\underline{\tilde{\varepsilon}\tilde{\varepsilon}}} = (P\,SNR_p)^{-1}\mathbf{I}$. This implies that the errors of estimation of the CIR samples are mutually uncorrelated, and ensures equal MSE distribution across all subcarriers as explained before.

Another property of interest is the cross-correlation between the estimate and the estimation error, which is expressed as

$$\begin{aligned}\sigma^2_{\tilde{\varepsilon}\tilde{h}} &= N^{-1}\text{E}[\hat{\mathbf{h}}^{\text{cls H}}\tilde{\boldsymbol{\varepsilon}}] \\ &= \text{tr}[\text{E}(\hat{\underline{\mathbf{h}}}^{\text{cls}}\hat{\underline{\mathbf{h}}}^{\text{cls H}}) - \text{E}(\underline{\mathbf{h}}\hat{\underline{\mathbf{h}}}^{\text{cls H}})] \\ &= \text{tr}(\mathbf{R}_{\underline{\hat{\mathbf{h}}\hat{\mathbf{h}}}} - \mathbf{R}_{\underline{\mathbf{hh}}}) \\ &= SNR_p^{-1}\text{tr}(\mathbf{S}) \\ &= (s\,SNR_p)^{-1} = \sigma^2_{\tilde{\varepsilon}}\end{aligned}. \tag{3.38}$$

It can be seen that this cross-correlation is always non-zero in the presence of noise and is equal to MSE.

3.3.4.2 Optimal CMMSE mode

In the optimal regime, both CIR correlation and SNR at the pilot subcarriers are considered to be either known or perfectly estimated at the OFDM receiver, i.e. $\bar{\mathbf{R}}_{\underline{\mathbf{hh}}} = \mathbf{R}_{\underline{\mathbf{hh}}}$ and $\bar{\sigma}^2_{w} = \sigma^2_{w}$. In this particular scenario, $\mathbf{R}_{\underline{\boldsymbol{\eta}\boldsymbol{\eta}}} = SNR_p^{-2}\mathbf{S}\mathbf{R}^{-1}_{\underline{\hat{\mathbf{h}}\hat{\mathbf{h}}}}\mathbf{S}$ in (3.33), and the estimator

$$\hat{\mathbf{h}}^{\text{cmmse}} = \mathbf{F}\mathbf{B}\hat{\underline{\mathbf{h}}}^{\text{cmmse}} = \sigma_p^{-2}\mathbf{F}\mathbf{B}(\mathbf{I} - SNR_p^{-1}\mathbf{S}\mathbf{R}^{-1}_{\underline{\hat{\mathbf{h}}\hat{\mathbf{h}}}})\mathbf{S}\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}^{\text{H}}\mathbf{X}^{\mathbf{p}\,\text{H}}\mathbf{y}^{\mathbf{p}} \tag{3.39}$$

efficiently achieves minimum MSE (MMSE) at its output, given by the following expression resultant from (3.34):

$$
\begin{aligned}
\sigma_{\varepsilon}^2 &= \mathrm{tr}[SNR_{\mathrm{p}}^{-1}\mathbf{S} - SNR_{\mathrm{p}}^{-2}\mathbf{S}(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\mathrm{p}}^{-1}\mathbf{S})^{-1}\mathbf{S}] \\
&= SNR_{\mathrm{p}}^{-1}\mathrm{tr}[\mathbf{R}_{\underline{\mathbf{hh}}}(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\mathrm{p}}^{-1}\mathbf{S})^{-1}\mathbf{S}] \\
&= \mathrm{tr}[\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2}(SNR_{\mathrm{p}}\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2} + \mathbf{I})^{-1}\mathbf{S}] \\
&= P\,\mathrm{tr}[\mathbf{U}_{\underline{\mathbf{hh}}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}(P\,SNR_{\mathrm{p}}\mathbf{U}_{\underline{\mathbf{hh}}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}} + \mathbf{I})^{-1}\mathbf{S}] \\
&= P\,\mathrm{tr}[\mathbf{S}\,\mathbf{U}_{\underline{\mathbf{hh}}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(P\,SNR_{\mathrm{p}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1}\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}]
\end{aligned}, \tag{3.40}
$$

where the identity $\mathbf{B} - \mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B} = \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}$ [102] is used, and EVD is applied to the positive-semidefinite Hermitean matrix $P^{-1}\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2}$, so that it could be factored as $P^{-1}\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2} = \mathbf{U}_{\underline{\mathbf{hh}}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}$, with the diagonal matrix $\mathbf{\Lambda}_{\underline{\mathbf{hh}}}$ containing eigenvalues λ_l, $l \in [0, L-1]$.

For the equispaced pilot pattern, the eigenvalues in $\mathbf{\Lambda}_{\underline{\mathbf{hh}}}$ and the columns of $\mathbf{U}_{\underline{\mathbf{hh}}}$ are correspondingly eigenvalues and eigenvectors of just the CIR correlation matrix $\mathbf{R}_{\underline{\mathbf{hh}}}$. Thus, the MMSE expression (3.40) becomes

$$
\sigma_{\varepsilon,\mathrm{es}}^2 = \mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(P\,SNR_{\mathrm{p}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1}] = \sum_{l=0}^{L-1}\frac{\lambda_l}{1 + P\,SNR_{\mathrm{p}}\lambda_l}. \tag{3.41}
$$

Henceforth we will term an eigenvalue λ_l as the *l*th *principal component*. Due to the rank constraint in (2.11), there are only $K \le L$ non-zero principal components. It should be noted that at the higher operational SNR_{p} values, the MSE difference between various λ_l distributions of L principal components becomes smaller due to dominant factor $P\,SNR_{\mathrm{p}}$ in (3.41).

To determine the worst-case performance of the channel estimator (3.39), it is important to analyse MMSE function (3.40) for maximum and minimum, depending on λ_l distribution.

The maximum of σ_{ε}^2 is derived in Appendix C.6 and is equal to

$$
\bar{\sigma}_{\varepsilon}^2 = (1 + s\,SNR_{\mathrm{p}})^{-1}. \tag{3.42}
$$

$\bar{\sigma}_{\varepsilon}^2$ (3.42) represents an *upper bound* on MMSE, such that for any λ_l distribution under the given SNR_{p}, $\sigma_{\varepsilon}^2 \le \bar{\sigma}_{\varepsilon}^2$. As shown in Appendix C.6, the MMSE function (3.40) attains $\bar{\sigma}_{\varepsilon}^2$ value when $\mathbf{\Lambda}_{\underline{\mathbf{hh}}} = s\,P^{-1}\mathbf{I}$, which corresponds to $\mathbf{R}_{\underline{\mathbf{hh}}} = s\,\mathbf{S}$. For the equispaced pilot arrangement, (3.42) is reduced to

$$
\bar{\sigma}_{\varepsilon,\mathrm{es}}^2 = L/(L + P\,SNR_{\mathrm{p}}), \tag{3.43}
$$

due to $s = P\,L^{-1}$.

Note that $\bar{\sigma}_{\varepsilon}^2$ does not depend on the path delay distribution, but is rather determined by the assumed CIR length *L* as the latter affects the value taken by s in (3.42). Therefore accurate establishment of the maximum excess delay of the channel is imperative. It is also important to point out that in general the ultimate case of $\mathbf{R}_{\underline{\mathbf{hh}}} = s\,\mathbf{S}$ is complex-valued, whereas $\mathbf{R}_{\underline{\mathbf{hh}}}$ (2.11) is known to have only real elements. Hence for the non-equispaced pilots it is always true that $\sigma_{\varepsilon}^2 < \bar{\sigma}_{\varepsilon}^2$.

In Appendix C.6 it is proved that the *lower bound* on MMSE, $\breve{\sigma}_{\varepsilon}^2 \le \sigma_{\varepsilon}^2$, is met when the channel is single-path with $\mathbf{R}_{\underline{\mathbf{hh}}}$ being of unity rank. For a general case, the closed-form $\breve{\sigma}_{\varepsilon}^2$ expression is difficult to derive, due to its dependence on the delay of the principal component. But for the equispaced pilot pattern, it follows directly from (3.41), irrespective of the excess delay of the propagation path:

$$\breve{\sigma}_{\varepsilon,\mathrm{es}}^2 = 1/(1 + P\,SNR_{\mathrm{p}})\,. \tag{3.44}$$

The physical interpretation of the bounds (3.43) and (3.44) is straightforward. The maximum MMSE, $\breve{\sigma}_{\varepsilon,\mathrm{es}}^2$, corresponds to the case, when all the principal components of the channel are equal, namely $\lambda_l = L^{-1}$, $\forall l \in [0, L-1]$, whereas the minimum MMSE, $\breve{\sigma}_{\varepsilon,\mathrm{es}}^2$, is inherent to the channel with only one principal component indexed l_0, $l_0 \in [0, L-1]$, i.e. $\lambda_l = \begin{cases} 1, & l = l_0 \\ 0, & l \ne l_0 \end{cases}$. In a special scenario, when the unity-ranked $\mathbf{R}_{\underline{\mathbf{hh}}}$ is diagonal, $\breve{\sigma}_{\varepsilon,\mathrm{es}}^2$ expresses MMSE of the flat fading channel response. An example of MMSE for various λ_l values is shown in Fig.3.3, where the 2-path channel is considered. One can see that the maximum is indeed located at $\lambda_0 = \lambda_1 = 1/2$ while the least MMSE values are observed when either λ_0 or λ_1 is close to zero.

Similar to the CLS mode, here we present the analysis of the properties inherent to the optimal CMMSE output. $\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}} = SNR_{\mathrm{p}}^{-1}\mathbf{R}_{\underline{\mathbf{hh}}}(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\mathrm{p}}^{-1}\mathbf{S})^{-1}\mathbf{S}$, defined through (3.33), (3.34) and (3.40), takes a diagonal form if the two following conditions hold simultaneously. First, the pilot pattern has to be equispaced, so that $\mathbf{S} = P^{-1}\mathbf{I}$. Second, $\mathbf{R}_{\underline{\mathbf{hh}}}$ has to be diagonal (e.g., this rule is true for the sample-spaced channels). Only subject to these two conditions, can one consider errors of estimation of the CIR samples as mutually uncorrelated. Apart from that, the property of equal MSE distribution across subcarriers holds if and only if $\mathbf{R}_{\underline{\mathbf{hh}}} = L^{-1}\mathbf{I}$, i.e. specifically for the channels with L equal principal components.

One can also show similar to derivation in Appendix C.4 that the cross-correlation between the estimate and the estimation error is equal to zero for the optimal CMMSE design:

$$\begin{aligned}
\sigma_{\underline{\varepsilon}\hat{\mathbf{h}}}^2 &= N^{-1}\mathrm{E}[\hat{\mathbf{h}}^{\mathrm{cmmse\,H}}\boldsymbol{\varepsilon}] \\
&= \mathrm{tr}[\mathrm{E}(\underline{\hat{\mathbf{h}}}^{\mathrm{cmmse}}\,\underline{\hat{\mathbf{h}}}^{\mathrm{cmmse\,H}}) - \mathrm{E}(\underline{\mathbf{h}}\,\underline{\hat{\mathbf{h}}}^{\mathrm{cmmse\,H}})] \\
&= \mathrm{tr}[\sigma_{\mathrm{p}}^4\mathbf{G}\mathbf{S}^{-1}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1}\mathbf{G}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^2\mathbf{G}\mathbf{S}^{-1}\mathbf{G}^{\mathrm{H}} - \sigma_{\mathrm{p}}^2\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1}\mathbf{G}^{\mathrm{H}}] \\
&= \mathrm{tr}[(\mathbf{I} - \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1})\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}(\mathbf{I} - \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1}\mathbf{S}) - \mathbf{R}_{\underline{\mathbf{hh}}}(\mathbf{I} - \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1}\mathbf{S})] \\
&= \mathrm{tr}[(\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}} - \mathbf{R}_{\underline{\mathbf{hh}}} - \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S})(\mathbf{I} - \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1}\mathbf{S})] = 0
\end{aligned}, \tag{3.45}$$

where $\mathbf{G} = \sigma_{\mathrm{p}}^{-2}(\mathbf{I} - \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1})\mathbf{S}$ and $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}} = \mathbf{R}_{\underline{\mathbf{hh}}} + \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}$. This result is well in accordance with the corollary of the Gauss-Markov theorem [39], and is in contrast to the CLS estimator, for which $\sigma_{\underline{\varepsilon}\hat{\mathbf{h}}}^2$ is equal to the estimation MSE.

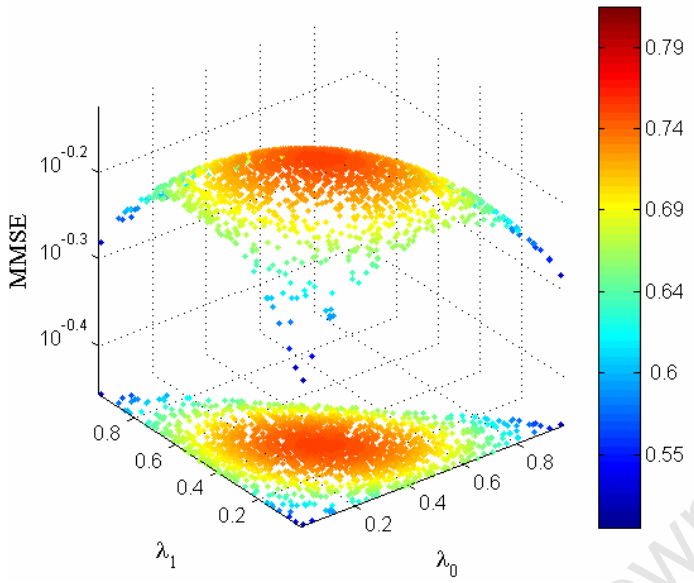


Fig.3.3. MMSE variation depending on distribution of the principal components in the 2-path channel estimation

3.3.4.3 CMMSE mode with imperfect noise variance information

In some cases information about the noise variance might be imperfect, i.e. $\overline{\sigma}_{\mathrm{w}}^2 \neq \sigma_{\mathrm{w}}^2$. It results in deviation of MSE from the optimal case (3.40) that is analysed in Appendix C.7. It is shown that the suboptimal CMMSE is no worse than the CLS estimator, i.e. $\sigma_{\varepsilon}^2 \leq \sigma_{\tilde{\varepsilon}}^2$, if it is designed for $\overline{\sigma}_{\mathrm{w}}^2$ satisfying $0 \leq \overline{\sigma}_{\mathrm{w}}^2 \leq 2\sigma_{\mathrm{w}}^2$. If the latter condition is not met, MSE experiences dramatic growth, proportional to $\overline{\sigma}_{\mathrm{w}}^2 - \sigma_{\mathrm{w}}^2$.

The main conclusion, following from $\overline{\sigma}_{\mathrm{w}}^2$ mismatch analysis (Appendix C.7), is a recommendation to introduce $\overline{\sigma}_{\mathrm{w}}^2$ ($\overline{SNR}_{\mathrm{p}}$) bias in the direction of $\overline{\sigma}_{\mathrm{w}}^2 < \sigma_{\mathrm{w}}^2$ ($\overline{SNR}_{\mathrm{p}} > SNR_{\mathrm{p}}$) if unbiased estimation of σ_{w}^2 cannot be realised due to some reason.

3.3.4.4 Diagonal CMMSE mode for equispaced pilot pattern

In comparison with CLS (3.8), the optimal CMMSE estimator (3.21), as well as its recursive counterpart (3.27) may be regarded as quite inefficient from the computational speed standpoint due to the necessity to acquire and invert a full-size ($L \times L$) CIR-plus-noise correlation matrix $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}$. In a particular implementation (e.g., MIMO with large dimensions), one might want to bring the complexity down to the CLS level, which is surprisingly low for the equispaced pilot pattern (Tab.3.2). The non-comb pilot case is not so interesting in this sense because the CLS and RMMSE estimators share the same complexity order. Here we show for the case of the equispaced pilot arrangement that it is possible to make a simplified suboptimal CMMSE implementation, which would not be worse than CLS mode in terms of the estimation MSE. This low-complexity mode is enabled by adopting the half-matched channel statistics for the estimator design, assuming that only the noise variance and power-delay profile (PDP) are known, so that $\overline{\mathbf{R}}_{\underline{\mathbf{hh}}} = \mathbf{R}_{\underline{\mathbf{hh}}} \circ \mathbf{I}_{L \times L}$ and $\overline{\sigma}_{\mathrm{w}}^2 = \sigma_{\mathrm{w}}^2$. The diagonal form of $\overline{\mathbf{R}}_{\underline{\mathbf{hh}}}$ stipulates this simplified estimator to be called

diagonal. By denoting PDP samples $g_l = [\mathbf{R}_{\underline{\mathbf{hh}}}]_{l,l}$, $l \in [0, L-1]$, the diagonal CMMSE and RMMSE schemes are correspondingly described as

$$\hat{\mathbf{h}}^{\text{dcmmse}} = \mathbf{F}\,\mathbf{B}\,\hat{\underline{\mathbf{h}}}^{\text{dcmmse}} = \sigma_{\text{p}}^{-2} P^{-1} \mathbf{F}\,\mathbf{B}\,[\mathbf{I} - \text{dg}(P\,SNR_{\text{p}}\mathbf{p} + \mathbf{1})^{-1}]\,\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}^{\text{H}}\mathbf{X}^{\mathbf{p}\,\text{H}}\mathbf{y}^{\mathbf{p}}, \tag{3.46}$$

$$\hat{\mathbf{h}}_m^{\text{drmmse}} = \mathbf{F}\,\mathbf{B}\,\hat{\underline{\mathbf{h}}}_m^{\text{drmmse}} = \sigma_{\text{p}}^{-2} P^{-1} \mathbf{F}\,\mathbf{B}\,[\mathbf{I} - \text{dg}(P\,SNR_{\text{p}}\hat{\tilde{\mathbf{p}}}_m)^{-1}]\,\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}^{\text{H}}\mathbf{X}^{\mathbf{p}\,\text{H}}\mathbf{y}_m^{\mathbf{p}}, \tag{3.47}$$

where $\mathbf{p} = [g_l]_{L\times 1}$ is PDP, $\mathbf{1}$ is the vector with all ones, and $\hat{\tilde{\mathbf{p}}}_m = (1-\alpha_{\text{FF}})\,\hat{\tilde{\mathbf{p}}}_{m-1} + \alpha_{\text{FF}}\,\text{dg}(\tilde{\underline{\mathbf{h}}}_m)^{\text{H}}\tilde{\underline{\mathbf{h}}}_m$ is the estimate of $\tilde{\mathbf{p}} = \text{dg}(\mathbf{R}_{\tilde{\underline{\mathbf{h}}}\tilde{\underline{\mathbf{h}}}})$.

Note that the estimators (3.46) and (3.47) are optimal if $\mathbf{R}_{\underline{\mathbf{hh}}}$ is diagonal, i.e. when CIR samples are independent CGRVs, as for example in the sample-spaced multipath channel models (refer to Subsection 2.2.2).

MSE of the diagonal CMMSE estimator is derived in Appendix C.8, with the resultant expression:

$$\sigma_{\varepsilon,\text{es}}^2 = \sum_{l=0}^{L-1} \frac{g_l}{1 + P\,SNR_{\text{p}}\,g_l}. \tag{3.48}$$

In analogy with the MMSE function (3.41), taking into account that $\text{tr}(\mathbf{R}_{\underline{\mathbf{hh}}}) = 1$, one can express the upper and lower bounds on $\sigma_{\varepsilon,\text{es}}^2$ (3.48) as $\bar{\sigma}_{\varepsilon,\text{es}}^2 = L/(L + P\,SNR_{\text{p}})$ and $\breve{\sigma}_{\varepsilon,\text{es}}^2 = 1/(1 + P\,SNR_{\text{p}})$, where $\bar{\sigma}_{\varepsilon,\text{es}}^2 < \breve{\sigma}_{\varepsilon,\text{es}}^2$ (3.15). However, unlike the eigenvalue decomposition identifying the independent principal components λ_l in case of the optimal CMMSE design, in most practical scenarios of the generally non-sample-spaced channels, $g_l > 0$ for $\forall l \in [0, L-1]$ due to the band limitation of CIR. In other words, for a random propagation environment, (3.48) would take a value closer to the upper bound, $\bar{\sigma}_{\varepsilon,\text{es}}^2$, with much higher probability than MMSE (3.41). Thus, performance of the suboptimal diagonal estimator is worse than that of the optimal CMMSE, except for the sample-spaced channels.

3.3.4.5 Robust CMMSE mode

Finally, let us consider the estimator design that does not require any correlation knowledge, but only the noise variance information ($\bar{\sigma}_{\text{w}}^2 = \sigma_{\text{w}}^2$). The subsequent analysis proves that setting $\overline{\mathbf{R}}_{\underline{\mathbf{hh}}} = s\mathbf{S}$, where $s = 1/\text{tr}(\mathbf{S})$, results in MSE performance equal to the upper MMSE bound, $\bar{\sigma}_{\varepsilon}^2$ (3.42), for any kind of the channel response statistics. Hence the corresponding estimator is called robust.

Calculating estimation MSE for an arbitrary CIR correlation, described by $\mathbf{R}_{\underline{\mathbf{hh}}}$, yields (Appendix C.9)

$$\sigma_{\varepsilon}^2 = (1 + s\,SNR_{\text{p}})^{-1} = \bar{\sigma}_{\varepsilon}^2. \tag{3.49}$$

It is important to note that for the higher SNRs, $\bar{\sigma}_{\varepsilon}^2 \approx s^{-1} SNR_{\text{p}}^{-1}$, i.e. performance of the robust CMMSE estimator is equivalent to that of the CLS mode. Thus, the robust design does not improve estimation accuracy in comparison with the simpler CLS algorithm.

3.3.5 Extension to MIMO

In analogy with the CLS estimator described in Section 3.2, CMMSE and RMMSE estimator can be implemented separately for each SL. Thus, for the MIMO configuration, formulas (3.21) and (3.27) become

$$\hat{\mathbf{h}}_m^{\text{cmmse}}(j,i) = \sigma_{\text{p}}^{-2}\mathbf{F}\mathbf{B}[\mathbf{I} - SNR_{\text{p}}^{-1}\mathbf{S}(i)(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\text{p}}^{-1}\mathbf{S}(i))^{-1}]\mathbf{S}(i)\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}(i)^{\text{H}}\mathbf{X}^{\mathbf{p}}(i)^{\text{H}}\mathbf{y}_m^{\mathbf{p}}(j,i), \tag{3.50}$$

$$\hat{\mathbf{h}}_m^{\text{rmmse}}(j,i) = \sigma_{\text{p}}^{-2}\mathbf{F}\mathbf{B}[\mathbf{I} - SNR_{\text{p}}^{-1}\mathbf{S}(i)\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(i)]\mathbf{S}(i)\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}(i)^{\text{H}}\mathbf{X}^{\mathbf{p}}(i)^{\text{H}}\mathbf{y}_m^{\mathbf{p}}(j,i), \tag{3.51}$$

where we assume SNR being the same at all the Rx antenna inputs.

Note that the weighting matrices $[\mathbf{I} - SNR_{\text{p}}^{-1}\mathbf{S}(i)(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\text{p}}^{-1}\mathbf{S}(i))^{-1}]\mathbf{S}(i)$ and $[\mathbf{I} - SNR_{\text{p}}^{-1}\mathbf{S}(i)\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(i)]\mathbf{S}(i)$ have to be computed only N_{tx} times, whereas the overall number of estimators is $N_{\text{tx}}N_{\text{rx}}$. Thus, the filtering part is of the same complexity for systems of any order of the receive diversity, irrespective of the algorithm used.

Based on the channel independence assumption with regard to different SLs, namely $\text{E}[\underline{\mathbf{h}}_m(j_1,i)\underline{\mathbf{h}}_m(j_2,i)^{\text{H}}] = \mathbf{0}$ for $j_1 \neq j_2$ (refer to Subsection 2.3.1), one can show that

$$\left(N_{\text{rx}}^{-1}\text{E}\left[\sum_{j=0}^{N_{\text{rx}}-1}\tilde{\underline{\mathbf{h}}}_m(j,i)\tilde{\underline{\mathbf{h}}}_m(j,i)^{\text{H}}\right]\right)^{-1} = N_{\text{rx}}\left(\sum_{j=0}^{N_{\text{rx}}-1}\text{E}[\underline{\mathbf{h}}_m(j,i)+\tilde{\underline{\boldsymbol{\varepsilon}}}_m(j,i)][\underline{\mathbf{h}}_m(j,i)+\tilde{\underline{\boldsymbol{\varepsilon}}}_m(j,i)]^{\text{H}}\right)^{-1} = \mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}^{-1}(i) = \mathbf{T}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}(i), \tag{3.52}$$

where $\tilde{\underline{\boldsymbol{\varepsilon}}}_m(j,i) = \sigma_{\text{p}}^{-2}\mathbf{S}(i)\mathbf{B}^{\text{H}}\mathbf{F}^{\text{H}}\mathbf{C}(i)^{\text{H}}\mathbf{X}^{\mathbf{p}}(i)^{\text{H}}\mathbf{w}_m^{\mathbf{p}}(j,i)$, $\mathbf{R}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}(i) = \mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\text{p}}^{-1}\mathbf{S}(i)$ (refer to (3.11) and (3.24)), and $\text{E}[\tilde{\underline{\boldsymbol{\varepsilon}}}_m(j_1,i)\tilde{\underline{\boldsymbol{\varepsilon}}}_m(j_2,i)^{\text{H}}] = \mathbf{0}$ for $j_1 \neq j_2$ due to the independent pilot pattern used, i.e. the inverse of expectation of the sum of instantaneous correlations, $\tilde{\underline{\mathbf{h}}}_m(j,i)\tilde{\underline{\mathbf{h}}}_m(j,i)^{\text{H}}$, corresponding to different SLs, associated with the given Tx antenna *i*, is equal to the precision matrix describing each SL individually. Thus, the estimate of the precision matrix $\mathbf{T}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}(i) = \left(\text{E}[\tilde{\underline{\mathbf{h}}}_m(j,i)\tilde{\underline{\mathbf{h}}}_m(j,i)^{\text{H}}]\right)^{-1}$ (3.26) can be computed using the updating term

$$\mathbf{V}_m(i) = \left(\sum_{j=0}^{N_{\text{rx}}-1}\tilde{\underline{\mathbf{h}}}_m(j,i)\tilde{\underline{\mathbf{h}}}_m(j,i)^{\text{H}}\right)\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m-1}(i), \tag{3.53}$$

yielding

$$\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(i) = \frac{1}{(1-\alpha_{\text{FF}})}\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m-1}(i)\left[\mathbf{I} - \frac{\mathbf{V}_m(i)}{N_{\text{rx}}(1-\alpha_{\text{FF}})/\alpha_{\text{FF}} + \text{tr}[\mathbf{V}_m(i)]}\right]. \tag{3.54}$$

Averaging of $\mathbf{T}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}}(i)$ across independent SLs can potentially provide faster convergence of the RMMSE algorithm to the steady state (CMMSE) in the MSE sense. In this regard higher receive diversity order would improve estimation accuracy.

In case of the equispaced pilot subcarriers, $\mathbf{S}(i) = \mathbf{S}(0) = P^{-1}\mathbf{I}$ and $\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(i) = \hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(0) = \hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}$. Hence the filtering part is the same for all the estimators and is dominated by the product with the updated matrix $(\mathbf{R}_{\underline{\mathbf{hh}}} + P^{-1}SNR_{\text{p}}^{-1}\mathbf{I})^{-1}$ in the CMMSE scheme (3.50), and $\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}$ in the RMMSE scheme (3.51).

3.4 Robust two-dimensional constrained estimator

The reader could have noticed so far that the channel estimators, described in Section 3.2 and 3.3, perform processing of the pilot observations separately for each block. In the case of the RMMSE algorithm, the recursive operation is inherent only to estimation of the CIR correlation, i.e. the statistics and not the actual response of the channel over a sequence of blocks. Hence all the previously presented estimators were termed as intrablock or one-dimensional (1D) and benefited from the deterministic and statistical frequency-selectivity properties, namely the finite CIR length and intrablock CIR correlation. On the other hand, the interblock CIR variation intensity was not taken into account.

The linear time-varying (LTV) channel model (Section 2.2) specifies the lowpass stationary stochastic Doppler process to govern fading with time. The bound on the maximum Doppler shift (2.41), stipulated by the reasonable system design, allows to regard this variation relatively slow. The inherent strong interblock channel response correlation, associated with a low-order model of the Doppler spectrum, could potentially be exploited in the estimation algorithm.

In this section, we introduce the two-dimensional (2D) estimator concept as an extension, which supplements the intrablock filtering with the interblock one. Unlike the intrablock CLS, CMMSE and RMMSE, 2D estimators can be regarded as the estimators with memory, which produce the estimate based on the current pilot block and a number of the preceding ones. We emphasise the realisable and reconfigurable suboptimal solutions rather than the optimal ones due to the challenges imposed by adaptation to the concrete propagation environment, which may generally be characterised as stationary-stochastic only during limited periods of time (e.g., in practical mobility scenarios).

3.4.1 Design and implementation of constrained LS-MMSE estimator

It has been established in Subsection 3.3.4, as a corollary to equation (3.36), that errors $\underline{\tilde{\boldsymbol{\varepsilon}}}_m = [\underline{\tilde{\varepsilon}}_{l,m}]_{L\times 1}$ at the output of the CLS estimator,

$$\underline{\tilde{\mathbf{h}}}_m = \sigma_{\mathrm{p}}^{-2}\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\,\mathrm{H}}\mathbf{y}_m^{\mathbf{p}} = \underline{\mathbf{h}}_m + \underline{\tilde{\boldsymbol{\varepsilon}}}_m, \tag{3.55}$$

represent CGRVs, uncorrelated with the CIR waveform $\underline{\mathbf{h}}_m$, that is true both inside a single block and across different blocks. Another important observation is that the error correlation matrix, $\mathbf{R}_{\underline{\tilde{\varepsilon}}\underline{\tilde{\varepsilon}}} = \mathrm{E}[\underline{\tilde{\boldsymbol{\varepsilon}}}_m \underline{\tilde{\boldsymbol{\varepsilon}}}_m^{\mathrm{H}}] = SNR_{\mathrm{p}}^{-1}\mathbf{S}$ (3.12), is independent of the channel response statistics. These two useful properties allow for optimal processing of the CLS output by another linear estimation algorithm.

The CLS estimates (3.55) can be smoothed by applying parallel Wiener filters to the *L* estimated CIR samples (elements of the vector $\underline{\tilde{\mathbf{h}}}_m$) [56][137]. These are the FIR filters, which improve accuracy of the estimates by taking into account the interblock CIR correlation due to the slow channel variation.

Consider the design of the Wiener filter for the *l*th branch (*l*th element of $\underline{\tilde{\mathbf{h}}}_m = [\underline{\tilde{h}}_{l,m}]_{L\times 1}$). We try to estimate $\underline{h}_{l,m}$ based on the present and a sequence of the past observations, i.e. processing the samples $\underline{\tilde{h}}_{l,m-M+1}$ to $\underline{\tilde{h}}_{l,m}$. Here M is regarded as the filter length that can be infinite in theory, but is always limited to a reasonable value by the computational complexity considerations.

Let the vector $\mathbf{r} = [r_\nu]_{M\times 1}$ denote the Doppler autocorrelation function (ACF), evaluated over the lag sequence $\nu \in [0, M-1]$ (refer to Subsection 2.2.3 for definition), $\mathbf{f}_l = [f_{l,\nu}]_{M\times 1}$ specify the set of the FIR coefficients, and $\underline{\tilde{\mathbf{h}}}_{l,m} = [\underline{\tilde{h}}_{l,k}]_{M\times 1}$, $k \in [m-M+1, m]$ store observation memory of the *l*th CIR sample after CLS processing. One can use the orthogonality principle [39] to find the optimal FIR coefficients $\mathbf{f}_l$. It says that the MMSE-optimal linear estimator of the CIR sample $\underline{h}_{l,m}$,

$$\underline{\hat{h}}_{l,m} = \mathbf{f}_l^{\mathrm{H}} \mathbf{J}\, \underline{\tilde{\mathbf{h}}}_{l,m}, \tag{3.56}$$

is obtained when the estimation error,

$$\underline{\varepsilon}_{l,m} = \underline{\hat{h}}_{l,m} - \underline{h}_{l,m}, \tag{3.57}$$

is orthogonal to each observation in the sequence $\underline{\tilde{h}}_{l,m-M+1}$ to $\underline{\tilde{h}}_{l,m}$, i.e.

$$\mathrm{E}[\underline{\varepsilon}_{l,m} \underline{\tilde{\mathbf{h}}}_{l,m}^{\mathrm{H}}] = \mathbf{0}_{1\times M}. \tag{3.58}$$

$\mathbf{J}$ in (3.56) is the exchange matrix [103] with ones on the main cross-diagonal and zeros elsewhere, which has a property of $\mathbf{J}\,\mathbf{J} = \mathbf{I}$.

Substitution of (3.56) into (3.57) and (3.58) yields

$$\mathbf{f}_l^{\mathrm{H}} \mathbf{J}\, \mathrm{E}[\underline{\tilde{\mathbf{h}}}_{l,m} \underline{\tilde{\mathbf{h}}}_{l,m}^{\mathrm{H}}] = \mathrm{E}[\underline{h}_{l,m} \underline{\mathbf{h}}_{l,m}^{\mathrm{H}}], \tag{3.59}$$

where $\underline{\mathbf{h}}_{l,m} = [\underline{h}_{l,k}]_{M\times 1}$, $k \in [m-M+1, m]$, and it can be shown that

$$\mathrm{E}[\underline{h}_{l,m} \underline{\mathbf{h}}_{l,m}^{\mathrm{H}}] = [\mathbf{R}_{\underline{\mathbf{hh}}}]_{l,l}\, \mathbf{r}^{\mathrm{H}} \mathbf{J}, \tag{3.60}$$

$$\mathrm{E}[\underline{\tilde{\mathbf{h}}}_{l,m} \underline{\tilde{\mathbf{h}}}_{l,m}^{\mathrm{H}}] = \mathrm{E}[\underline{\mathbf{h}}_{l,m} \underline{\mathbf{h}}_{l,m}^{\mathrm{H}}] + \sigma_{\mathrm{w}}^2 \sigma_{\mathrm{p}}^{-2} [\mathbf{S}]_{l,l}\, \mathbf{I} \tag{3.61}$$

and

$$\mathrm{E}[\underline{\mathbf{h}}_{l,m} \underline{\mathbf{h}}_{l,m}^{\mathrm{H}}] = [\mathbf{R}_{\underline{\mathbf{hh}}}]_{l,l}\, \mathrm{toe}[\mathbf{r}, \mathbf{r}^{\mathrm{H}}] \tag{3.62}$$

($\mathrm{toe}[\mathbf{r}, \mathbf{r}^{\mathrm{H}}]$ denotes a square Toeplitz matrix with the first column $\mathbf{r}$). By letting the following notations for the *l*th PDP sample, the Doppler correlation matrix and the noise variance at the Wiener filter input, respectively:

$$g_l = [\mathbf{R}_{\underline{\mathbf{hh}}}]_{l,l}, \tag{3.63}$$

$$\mathbf{R} = (\text{toe}[\mathbf{r}, \mathbf{r}^{\mathrm{H}}])^{\mathrm{T}} = \mathbf{J}\,\text{toe}[\mathbf{r}, \mathbf{r}^{\mathrm{H}}]\mathbf{J}\,, \tag{3.64}$$

$$\rho_l = SNR_{\mathrm{p}}^{-1}[\mathbf{S}]_{l,l}\,, \tag{3.65}$$

the optimal set of the filter coefficients can be expressed as

$$\mathbf{f}_l^{\mathrm{H}} = g_l \mathbf{r}^{\mathrm{H}}\mathbf{J}(g_l \mathbf{J}\mathbf{R}\mathbf{J} + \rho_l \mathbf{I})^{-1}\mathbf{J}\,. \tag{3.66}$$

Hence (3.66) is reduced to

$$\mathbf{f}_l = g_l (g_l \mathbf{R} + \rho_l \mathbf{I})^{-1}\mathbf{r}\,. \tag{3.67}$$

Applying EVD $\mathbf{R} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}}$, the filter (3.67) attains the low-complexity form, which does not require a matrix inversion every time g_l or ρ_l statistics change:

$$\mathbf{f}_l = g_l \mathbf{U}(g_l \boldsymbol{\Lambda} + \rho_l \mathbf{I})^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{r} = g_l \sum_{k=0}^{M-1} \frac{w_k}{(g_l \lambda_k + \rho_l)}\mathbf{u}_k\,, \tag{3.68}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix containing the eigenvalues of $\mathbf{R}$, $\lambda_k = [\boldsymbol{\Lambda}]_{k,k}$, $k \in [0, M-1]$; the columns of $\mathbf{U}$, $\mathbf{u}_k = (\mathbf{U})_k$, are the eigenvectors, and $w_k = \mathbf{u}_k^{\mathrm{H}}\mathbf{r}$.

The overall output of the 2D CIR estimator, henceforth termed *CLS-MMSE*, can be described by the equation

$$\hat{\underline{\mathbf{h}}}_m^{\text{cls-mmse}} = \sum_{k=0}^{M-1} \boldsymbol{\Phi}_k \tilde{\underline{\mathbf{h}}}_{m-k}\,, \tag{3.69}$$

where $\boldsymbol{\Phi}_k = \text{dg}\left([f_{0,k} \quad \cdots \quad f_{L-1,k}]^{\mathrm{T}}\right)$.

The estimated CIR (3.69) is subsequently transformed into CFR to be used in the detector:

$$\hat{\mathbf{h}}_m^{\text{cls-mmse}} = \mathbf{F}\mathbf{B}\hat{\underline{\mathbf{h}}}_m^{\text{cls-mmse}}\,. \tag{3.70}$$

3.4.2 MSE analysis of constrained LS-MMSE estimator

Depending on the filterbank memory M, there are two cases to consider: *finite* and *infinite*. MSE analysis of the purely theoretical infinite-length filter will appear helpful to derive the robust design criteria for the Wiener filterbank and to determine the worst-case performance of the estimator.

3.4.2.1 MSE of estimator based on infinite Wiener filters

In derivation of the error variance at the output of the estimator (3.69), we consider the ideal filters first, for which $M = \infty$. In such a case the output of the filter, applied to the *l*th CIR sample, is described as

$$\hat{\underline{h}}_{l,m} = \sum_{k=0}^{\infty} f_{l,k}\,\tilde{\underline{h}}_{l,m-k}\,. \tag{3.71}$$

Revisiting the orthogonality principle (3.58), one can derive the MMSE-optimal FIR coefficients $f_{l,k}$ from the equation

$$\mathrm{E}[\underline{\varepsilon}_{l,m}\,\underline{\tilde{h}}^{*}_{l,m-\nu}] = 0\,, \tag{3.72}$$

where $\nu \in [0,\infty)$.

The transfer function, corresponding to the discrete-time sequence $f_{l,k}$, is derived in Appendix C.10 as

$$\Phi_l(\omega) = 1 - \frac{\rho_l}{\mu_l Q_l(\omega)}\,, \tag{3.73}$$

where $\omega \in [-\pi,\pi)$ denotes the normalised angular frequency,

$$\mu_l = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln \tilde{S}_l(\omega)\,\mathrm{d}\omega\right), \tag{3.74}$$

$$\tilde{S}_l(\omega) = S_l(\omega) + \rho_l\,, \tag{3.75}$$

$$S_l(\omega) = g_l S(\omega)\,, \tag{3.76}$$

and $Q_l(\omega)$ is found from the spectral factorisation

$$\tilde{S}_l(\omega) = \mu_l\, Q_l(\omega) Q_l^{*}(\omega)\,. \tag{3.77}$$

It is shown in Appendix C.11 that the variance of the estimation error, $\underline{\varepsilon}_{l,m}$ (3.57), at the output of the *l*th filter is equal to

$$\sigma^2_{\underline{\varepsilon}l} = g_l + \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_l(\omega)\Phi_l^{*}(\omega)\,\tilde{S}_l(\omega)\,\mathrm{d}\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_l(\omega)\,S_l(\omega)\,\mathrm{d}\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_l^{*}(\omega)\,S_l(\omega)\,\mathrm{d}\omega\,. \tag{3.78}$$

The filters (3.67) and (3.73) are optimal in the sense of the minimum MSE at the output only if the channel PDP g_l, the Doppler power spectral density (PSD) $S(\omega)$, and the noise variance σ^2_{w}, underlying the terms $\mathbf{r}$, ρ_l and μ_l, are known at the receiver. Unlike g_l, which can be estimated adaptively using CSI from sequential blocks (as well as from different SLs in the MIMO configuration) and σ^2_{w}, which can be measured on the virtual subcarriers (refer to Subsection 3.3.2 for additional details), accurate estimation of the Doppler PSD is highly impractical due to the necessity to transmit a very long sequence of blocks to capture the slow time variation of the channel response. Thus, the robust design of the Wiener filterbank necessitates such $S(\omega)$ setting, which would not lead to the variance of the estimation error, $\underline{\varepsilon}_{l,m}$, at the output of the *l*th filter being higher than the noise variance, ρ_l, at its input for $\forall l \in [0, L-1]$, i.e. which would ensure the stability of the estimator.

In Appendix C.12, it is shown that letting the bandlimited *flat* Doppler PSD,

$$\bar{S}(\omega) = \begin{cases} \pi/\tilde{\omega}_{\mathrm{D}}\,, & |\omega| \le \tilde{\omega}_{\mathrm{D}} \\ 0, & \text{otherwise} \end{cases}\,, \tag{3.79}$$

to compute the transfer function (3.73), ensures constant error variance at the output of the Wiener filter that is expressed as

$$\sigma^2_{\underline{\varepsilon}l} = \rho_l\left[1 - \left(g_l \rho_l^{-1}\Omega + 1\right)^{-\Omega^{-1}}\right], \tag{3.80}$$

where

$$\Omega = \pi / \tilde{\omega}_{\mathrm{D}} . \tag{3.81}$$

Here parameter $\tilde{\omega}_{\mathrm{D}}$ is different from its channel-model analogue in (2.28). It denotes the maximum normalised angular Doppler frequency as seen by the filter when not all the blocks carry pilot symbols. Let two sequential blocks with both pilot and data subcarriers be separated in time by $\underline{P}-1$ blocks, which transmit only data symbols. Then we define the *pilot periodicity* coefficient $\underline{P}$. The term "periodicity" is used intentionally instead of the term "spacing", to distinguish between the interblock and intrablock pilot arrangement. Decimation of the pilot block sequence by the factor $\underline{P}$ is equivalent to the Doppler spectrum broadening due to the total filter bandwidth reduction. This condition extends the maximum Doppler shift to

$$\tilde{\omega}_{\mathrm{D}} = \underline{P}\, \omega_{\mathrm{D}} . \tag{3.82}$$

The channel response in between the successive pilot-bearing blocks can either be assumed invariant or interpolated. The comprehensive overview of the applicable interpolation techniques can be found in the work [61]. We, however, do not address design of the optimal interblock interpolator in this thesis as it is a separate problem with regard to filtering of the pilot measurements. In fact, the interpolator can be incorporated after any CIR estimator described in this chapter, with no difference of the estimation algorithm used.

Given the variance of error at the output of the *l*th filter (3.80), the overall CFR estimation MSE is computed as

$$\sigma_{\varepsilon}^2 = \sum_{l=0}^{L-1} \sigma_{\underline{\varepsilon} l}^2 = \sum_{l=0}^{L-1} \rho_l \left[1 - \left(g_l \rho_l^{-1} \Omega + 1 \right)^{-\Omega^{-1}} \right]. \tag{3.83}$$

We also let $c_l = [\mathbf{S}]_{l,l}$ in (3.65) for a shorter notation, i.e. $\rho_l = SNR_{\mathrm{p}}^{-1} c_l$.

To determine the worst-case performance of the CLS-MMSE estimator (3.70) with the robust Doppler PSD design, it is necessary to analyse the MSE function (3.83) for maximum, depending on g_l distribution (channel PDP). In Appendix C.13, it is shown that MSE is upper-bounded by

$$\bar{\sigma}_{\varepsilon}^2 \approx \frac{1}{SNR_{\mathrm{p}}} \left[\frac{1}{s} - \frac{L}{P} \left(\frac{\Omega P SNR_{\mathrm{p}}}{L} + 1 \right)^{-\Omega^{-1}} \right], \tag{3.84}$$

where $s = \left(\sum_{l=0}^{L-1} c_l \right)^{-1}$ in accordance with (3.14).

The actual MSE (3.83) approaches the upper bound (3.84) when the channel PDP samples are given as

$$g_l = \left(\sum_{l=0}^{L-1} c_l^{\Omega+1} \right)^{-1} c_l^{\Omega+1} . \tag{3.85}$$

The exact magnitude of (3.84) is attained only for the comb pilot pattern, i.e. when $c_l = P^{-1}$, that in turn corresponds to the uniform PDP $g_l = L^{-1}$:

$$\breve{\sigma}^2_{\varepsilon,\mathrm{es}} = \frac{L}{P\,SNR_{\mathrm{p}}}\left[1-\left(\frac{\Omega\,P\,SNR_{\mathrm{p}}}{L}+1\right)^{-\Omega^{-1}}\right]. \tag{3.86}$$

It is also shown in Appendix C.13 that the robust CLS-MMSE estimator achieves the best accuracy in flat fading channels. The corresponding MSE, determined for the most precise filter in the Wiener filterbank, is regarded as the lower bound and is expressed as follows:

$$\breve{\sigma}^2_{\varepsilon} = SNR_{\mathrm{p}}^{-1} c_{l_0}\left[1-\left(\Omega\,SNR_{\mathrm{p}} c_{l_0}^{-1}+1\right)^{-\Omega^{-1}}\right], \tag{3.87}$$

where $l_0 = \arg\min_{l\in[0,L-1]} c_l$ identifies the excess delay, which matches the filter with the least output error variance.

In case of the equispaced pilot arrangement, all L filters have identical accuracy, and (3.87) becomes

$$\breve{\sigma}^2_{\varepsilon,\mathrm{es}} = (P\,SNR_{\mathrm{p}})^{-1}\left[1-\left(\Omega\,P\,SNR_{\mathrm{p}}+1\right)^{-\Omega^{-1}}\right]. \tag{3.88}$$

It should be noted that in contrast to the intrablock CMMSE estimator, for which the lower MSE bound is met in the case of an arbitrary excess delay of the sole principal component (refer to Subsection 3.4.2), the lower MSE bound of the CLS-MMSE estimator is more specific, being inherent to the sample-spaced excess delay only. The comb pilot pattern is interesting in the sense of identical statistics of the sample-spaced multipath channel to yield the lower and upper MSE bounds for both CMMSE and CLS-MMSE estimators.

3.4.2.2 MSE of estimator based on finite Wiener filters

From the theoretical standpoint, the finite-length filters have an obvious drawback in comparison with the infinite counterpart as there is no direct way to conduct spectral analysis of the MSE performance and establish the robust design conditions. Using numerical examples, we show here that the flat-shaped Doppler spectrum assumption yields robust design for filters of finite length as well. The analytical MSE derivation is presented first.

Let the set of FIR coefficients, corresponding to the flat Doppler spectrum bandlimited in the interval $\omega\in[-\overline{\omega}_{\mathrm{D}},\overline{\omega}_{\mathrm{D}})$, $\tilde{\omega}_{\mathrm{D}} \le \overline{\omega}_{\mathrm{D}} < \pi$, be denoted in analogy with (3.67) as

$$\bar{\mathbf{f}}_l = g_l(g_l\overline{\mathbf{R}}+\rho_l\mathbf{I})^{-1}\bar{\mathbf{r}}, \tag{3.89}$$

where the elements of the ACF vector $\bar{\mathbf{r}} = [\bar{r}_\nu]_{M\times 1}$, $\nu\in[0,M-1]$ are given as

$$\bar{r}_\nu = \sin(\overline{\omega}_{\mathrm{D}}\nu)/(\overline{\omega}_{\mathrm{D}}\nu), \tag{3.90}$$

and $\overline{\mathbf{R}} = \mathbf{J}\,\mathrm{toe}[\bar{\mathbf{r}},\bar{\mathbf{r}}^{\mathrm{H}}]\mathbf{J} = (\mathrm{toe}[\bar{\mathbf{r}},\bar{\mathbf{r}}^{\mathrm{H}}])^{\mathrm{T}}$ similar to (3.64). Note that setting $\overline{\omega}_{\mathrm{D}} \ge \tilde{\omega}_{\mathrm{D}}$ is important if the actual Doppler shift of the input process, $\tilde{\omega}_{\mathrm{D}}$, is not known precisely.

Then the (l_1,l_2) th element of the intrablock CIR error correlation matrix at the output of the CLS-MMSE estimator (3.69), based on the finite Wiener filters, is determined as follows:

$$\begin{aligned}
[\mathbf{R}_{\underline{\underline{\boldsymbol{\varepsilon}}}}]_{l_1,l_2} &= \mathrm{E}[\underline{\varepsilon}_{l_1,m}\underline{\varepsilon}^{*}_{l_2,m}] \\
&= \mathrm{E}[\underline{\hat{h}}_{l_1,m}\underline{\hat{h}}^{*}_{l_2,m}] + \mathrm{E}[\underline{h}_{l_1,m}\underline{h}^{*}_{l_2,m}] - \mathrm{E}[\underline{\hat{h}}_{l_1,m}\underline{h}^{*}_{l_2,m}] - \mathrm{E}[\underline{h}_{l_1,m}\underline{\hat{h}}^{*}_{l_2,m}] \\
&= \bar{\mathbf{f}}^{\mathrm{H}}_{l_1}\mathbf{J}\,\mathrm{E}[\underline{\underline{\tilde{\mathbf{h}}}}_{l_1,m}\underline{\underline{\tilde{\mathbf{h}}}}^{\mathrm{H}}_{l_2,m}]\mathbf{J}\bar{\mathbf{f}}_{l_2} + [\mathbf{R}_{\underline{\underline{\mathbf{hh}}}}]_{l_1,l_2} - \bar{\mathbf{f}}^{\mathrm{H}}_{l_1}\mathbf{J}\,\mathrm{E}[\underline{\underline{\tilde{\mathbf{h}}}}_{l_1,m}\underline{h}^{*}_{l_2,m}] - \mathrm{E}[\underline{h}_{l_1,m}\underline{\underline{\tilde{\mathbf{h}}}}^{\mathrm{H}}_{l_2,m}]\mathbf{J}\bar{\mathbf{f}}_{l_2},
\end{aligned} \tag{3.91}$$

where $l_1, l_2 \in [0, L-1]$, and the definitions (3.56)-(3.57) are used. Introducing notations:

$$g_{l_1 l_2} = [\mathbf{R}_{\underline{\mathbf{hh}}}]_{l_1, l_2}, \tag{3.92}$$

$$\rho_{l_1 l_2} = [\mathbf{R}_{\underline{\tilde{\varepsilon}}\underline{\tilde{\varepsilon}}}]_{l_1, l_2} = SNR_{\mathrm{p}}^{-1}[\mathbf{S}]_{l_1, l_2}, \tag{3.93}$$

with $g_{l_1 l_2} = g_{l_1}$ (3.63) and $\rho_{l_1 l_2} = \rho_{l_1}$ (3.65) if $l_2 = l_1$, and using (3.60)-(3.62), one can express (3.91) as

$$\begin{aligned}[\mathbf{R}_{\underline{\varepsilon}\underline{\varepsilon}}]_{l_1, l_2} &= g_{l_1 l_2} + \bar{\mathbf{f}}_{l_1}^{\mathrm{H}} \mathbf{J} (g_{l_1 l_2} \mathbf{J}\mathbf{R}\mathbf{J} + \rho_{l_1 l_2} \mathbf{I}) \mathbf{J} \bar{\mathbf{f}}_{l_2} - g_{l_1 l_2} \bar{\mathbf{f}}_{l_1}^{\mathrm{H}} \mathbf{r} - g_{l_1 l_2} \mathbf{r}^{\mathrm{H}} \bar{\mathbf{f}}_{l_2} \\ &= g_{l_1 l_2} + g_{l_1} g_{l_2} \bar{\mathbf{r}}^{\mathrm{H}} (g_{l_1} \bar{\mathbf{R}} + \rho_{l_1} \mathbf{I})^{-1} (g_{l_1 l_2} \mathbf{R} + \rho_{l_1 l_2} \mathbf{I}) (g_{l_2} \bar{\mathbf{R}} + \rho_{l_2} \mathbf{I})^{-1} \bar{\mathbf{r}} \\ &\quad - g_{l_1 l_2} g_{l_1} \bar{\mathbf{r}}^{\mathrm{H}} (g_{l_1} \bar{\mathbf{R}} + \rho_{l_1} \mathbf{I})^{-1} \mathbf{r} - g_{l_1 l_2} g_{l_2} \mathbf{r}^{\mathrm{H}} (g_{l_2} \bar{\mathbf{R}} + \rho_{l_2} \mathbf{I})^{-1} \bar{\mathbf{r}} \end{aligned} . \tag{3.94}$$

Hence one can compute MSE at the output of the lth filterbank branch:

$$\sigma_{\underline{\varepsilon} l}^2 = [\mathbf{R}_{\underline{\varepsilon}\underline{\varepsilon}}]_{l,l} = g_l \left[1 + g_l \bar{\mathbf{r}}^{\mathrm{H}} (g_l \bar{\mathbf{R}} + \rho_l \mathbf{I})^{-1} [(g_l \mathbf{R} + \rho_l \mathbf{I})(g_l \bar{\mathbf{R}} + \rho_l \mathbf{I})^{-1} \bar{\mathbf{r}} - 2\mathbf{r}] \right]. \tag{3.95}$$

If $\mathbf{r} = \bar{\mathbf{r}}$, (3.95) is reduced to

$$\bar{\sigma}_{\underline{\varepsilon} l}^2 = g_l \left[1 - g_l \bar{\mathbf{r}}^{\mathrm{H}} (g_l \bar{\mathbf{R}} + \rho_l \mathbf{I})^{-1} \bar{\mathbf{r}} \right] = g_l \left[1 - \bar{\mathbf{r}}^{\mathrm{H}} (\bar{\mathbf{R}} + \rho_l / g_l \mathbf{I})^{-1} \bar{\mathbf{r}} \right], \tag{3.96}$$

where the second expression can be used if $g_l \neq 0$.

Fig.3.4, Fig.3.5 and Fig.3.6 show the normalised MSE (NMSE) of the filter output, $\sigma_{\underline{\varepsilon} l}^2 / g_l$, for the various values of parameters M, $\tilde{\omega}_{\mathrm{D}}$, $\mathbf{r}$ and g_l / ρ_l when $\bar{\omega}_{\mathrm{D}} = 0.025\pi$ (2.41). We consider four different Doppler models, which have practical significance and are described in detail in Subsection 2.3.4 and Fig.2.9.

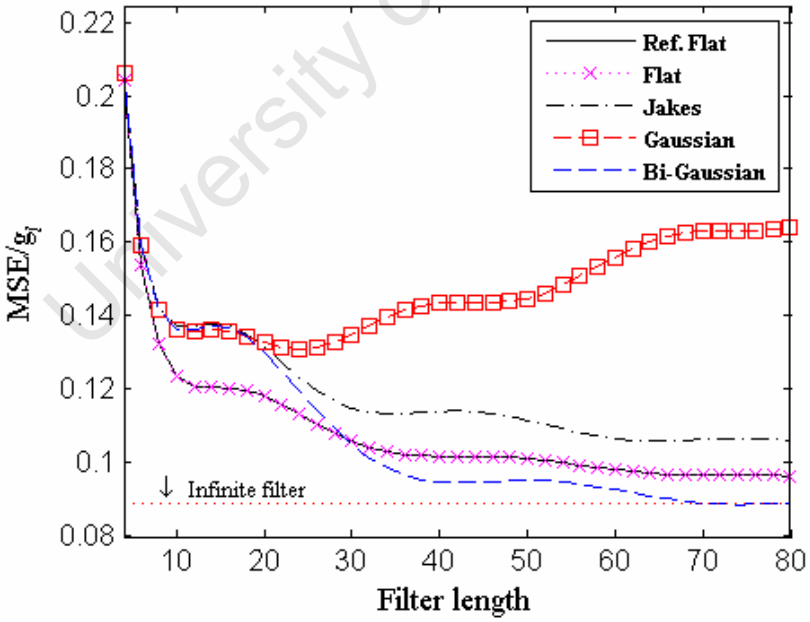


Fig.3.4. NMSE of the filter output for $\tilde{\omega}_{\mathrm{D}} = 0.025\pi$ and $g_l / \rho_l = 0\,\mathrm{dB}$

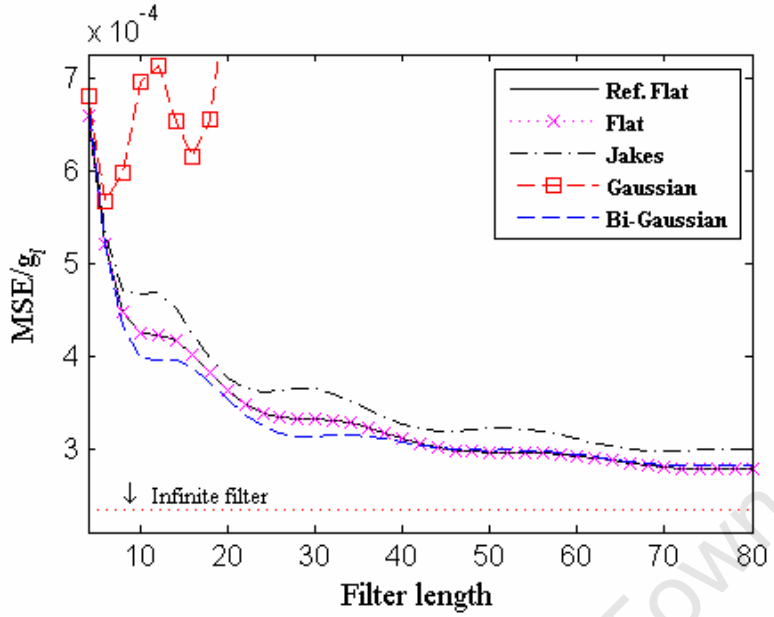


Fig.3.5. NMSE of the filter output for $\tilde{\omega}_{\mathrm{D}} = 0.025\pi$ and $g_l/\rho_l = 30\mathrm{dB}$

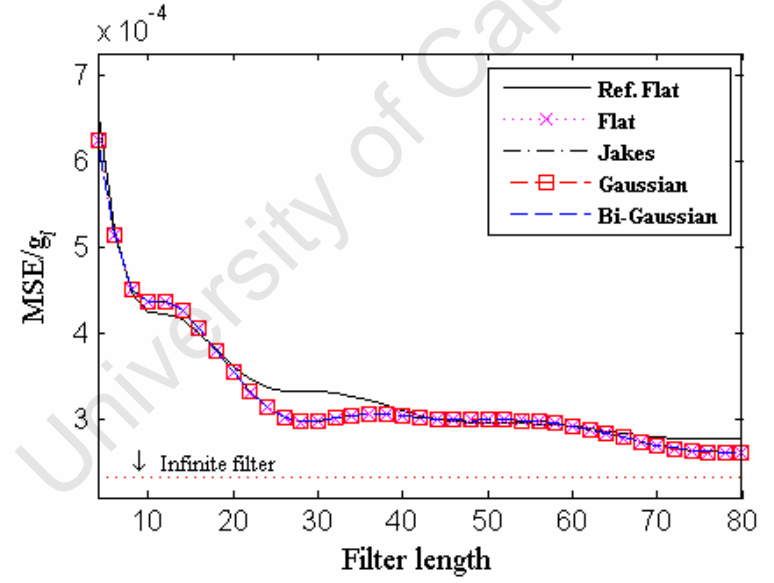


Fig.3.6. NMSE of the filter output for $\tilde{\omega}_{\mathrm{D}} = 0.0025\pi$ and $g_l/\rho_l = 30\mathrm{dB}$ (reference flat-spectrum model of the finite and infinite filter is designed for $\overline{\omega}_{\mathrm{D}} = 0.025\pi$)

One can see that the obligatory requirement of the robust filter design is an accurate prediction of the frequency boundaries of the Doppler spectrum. In Fig.3.4 and Fig.3.5, the Gaussian model extends over the filter's processing band that results in a dramatic MSE increase, which is most brightly manifested when the input noise variance is lower.

As the filter length M increases, MSE tends asymptotically towards the theoretical lower bound achieved by the infinite filter. However, for lower complexity schemes (small M) MSE can still be significantly higher than the lower bound.

All the graphs show that the shape of the Doppler spectrum does not noticeably influence MSE when M is large enough. For smaller M values, one can see MSE fluctuation around the predicted value (3.96) that decreases with lesser input noise power. Hence MSE of the filter output may always be regarded as fixed and approximately equal to (3.96), irrespective of the exact Doppler statistics.

Similar to (3.83), the overall CFR estimation MSE for the case of the robust finite-length filterbank is computed as

$$\begin{aligned}\sigma_{\varepsilon}^2 &= \sum_{l=0}^{L-1} \sigma_{\underline{\varepsilon} l}^2 \\ &= \sum_{l=0}^{L-1} g_l \left[1 + g_l \overline{\mathbf{r}}^{\mathrm{H}} (g_l \overline{\mathbf{R}} + \rho_l \mathbf{I})^{-1} [(g_l \mathbf{R} + \rho_l \mathbf{I})(g_l \overline{\mathbf{R}} + \rho_l \mathbf{I})^{-1} \overline{\mathbf{r}} - 2\mathbf{r}]\right] \\ &\approx \sum_{l=0}^{L-1} g_l \left[1 - g_l \overline{\mathbf{r}}^{\mathrm{H}} (g_l \overline{\mathbf{R}} + \rho_l \mathbf{I})^{-1} \overline{\mathbf{r}}\right] \end{aligned} . \tag{3.97}$$

Finally, it follows by induction that the lower and upper bounds on MSE (3.97) hold for the same PDP conditions (g_l distribution) as in case of the infinite filterbank.

3.4.3 Design and implementation of constrained MMSE-MMSE estimator

The robustly designed CLS-MMSE estimator has a very useful property, being characterised by a fixed output MSE, irrespective of the actual Doppler statistics. However, the estimation error is in general correlated with the channel response. The latter circumstance makes further optimal linear filtering of the observed CIR waveform impossible. Nevertheless, a suboptimal approach can be undertaken, based on the assumption of the uncorrelatedness of CIR and estimation error. Here we show that the CLS-MMSE estimator can be supplemented with the smoothing module relying on the CIR intrablock correlation that is similar to the CMMSE concept described in Subsection 3.3.1.

Given the error correlation matrix $\mathbf{R}_{\underline{\varepsilon}\underline{\varepsilon}}$ (3.94), inherent to the CLS-MMSE estimator $\hat{\underline{\mathbf{h}}}_m$ (3.69), the output of the proposed smoothing module can be described in analogy with (3.23)-(3.24) as

$$\hat{\underline{\mathbf{h}}}_m^{\text{cmmse-mmse}} = [\mathbf{I} - \mathbf{R}_{\underline{\varepsilon}\underline{\varepsilon}} (\mathbf{R}_{\underline{\mathbf{h}}\underline{\mathbf{h}}} + \mathbf{R}_{\underline{\varepsilon}\underline{\varepsilon}})^{-1}] \hat{\underline{\mathbf{h}}}_m^{\text{cls-mmse}} . \tag{3.98}$$

Henceforth we will refer to (3.98) as the *CMMSE-MMSE* CIR estimator. The corresponding CFR estimate is computed analogously to (3.70):

$$\hat{\mathbf{h}}_m^{\text{cmmse-mmse}} = \mathbf{F}\mathbf{B}\hat{\underline{\mathbf{h}}}_m^{\text{cmmse-mmse}} . \tag{3.99}$$

Using the robust design property of $\mathbf{R}_{\underline{\varepsilon}\underline{\varepsilon}}$ being the same for any Doppler process, (3.94) can be simplified to

$$
\begin{aligned}
[\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}]_{l_1,l_2} &= g_{l_1l_2} + g_{l_1}g_{l_2}\bar{\mathbf{r}}^{\mathrm{H}}(g_{l_1}\overline{\mathbf{R}}+\rho_{l_1}\mathbf{I})^{-1}(g_{l_1l_2}\overline{\mathbf{R}}+\rho_{l_1l_2}\mathbf{I})(g_{l_2}\overline{\mathbf{R}}+\rho_{l_2}\mathbf{I})^{-1}\bar{\mathbf{r}} \\
&\quad - g_{l_1l_2}g_{l_1}\bar{\mathbf{r}}^{\mathrm{H}}(g_{l_1}\overline{\mathbf{R}}+\rho_{l_1}\mathbf{I})^{-1}\bar{\mathbf{r}} - g_{l_1l_2}g_{l_2}\bar{\mathbf{r}}^{\mathrm{H}}(g_{l_2}\overline{\mathbf{R}}+\rho_{l_2}\mathbf{I})^{-1}\bar{\mathbf{r}} \\
&= g_{l_1l_2} + g_{l_1}g_{l_2}\bar{\mathbf{r}}^{\mathrm{H}}\overline{\mathbf{U}}(g_{l_1}\overline{\boldsymbol{\Lambda}}+\rho_{l_1}\mathbf{I})^{-1}(g_{l_1l_2}\overline{\boldsymbol{\Lambda}}+\rho_{l_1l_2}\mathbf{I})(g_{l_2}\overline{\boldsymbol{\Lambda}}+\rho_{l_2}\mathbf{I})^{-1}\overline{\mathbf{U}}^{\mathrm{H}}\bar{\mathbf{r}} \\
&\quad - g_{l_1l_2}g_{l_1}\bar{\mathbf{r}}^{\mathrm{H}}\overline{\mathbf{U}}(g_{l_1}\overline{\boldsymbol{\Lambda}}+\rho_{l_1}\mathbf{I})^{-1}\overline{\mathbf{U}}^{\mathrm{H}}\bar{\mathbf{r}} - g_{l_1l_2}g_{l_2}\bar{\mathbf{r}}^{\mathrm{H}}\overline{\mathbf{U}}(g_{l_2}\overline{\boldsymbol{\Lambda}}+\rho_{l_2}\mathbf{I})^{-1}\overline{\mathbf{U}}^{\mathrm{H}}\bar{\mathbf{r}} \\
&= g_{l_1l_2} + \mathrm{tr}[\overline{\mathbf{Z}}(g_{l_1}g_{l_2}(g_{l_1}\overline{\boldsymbol{\Lambda}}+\rho_{l_1}\mathbf{I})^{-1}(g_{l_1l_2}\overline{\boldsymbol{\Lambda}}+\rho_{l_1l_2}\mathbf{I})(g_{l_2}\overline{\boldsymbol{\Lambda}}+\rho_{l_2}\mathbf{I})^{-1} \\
&\quad - g_{l_1l_2}g_{l_1}(g_{l_1}\overline{\boldsymbol{\Lambda}}+\rho_{l_1}\mathbf{I})^{-1} - g_{l_1l_2}g_{l_2}(g_{l_2}\overline{\boldsymbol{\Lambda}}+\rho_{l_2}\mathbf{I})^{-1})] \qquad , \qquad (3.100)
\end{aligned}
$$

where $\overline{\mathbf{R}}$ is eigen-decomposed as $\overline{\mathbf{R}} = \overline{\mathbf{U}}\,\overline{\boldsymbol{\Lambda}}\,\overline{\mathbf{U}}^{\mathrm{H}}$ and $\overline{\mathbf{Z}} = \overline{\mathbf{U}}^{\mathrm{H}}\bar{\mathbf{r}}\,\bar{\mathbf{r}}^{\mathrm{H}}\overline{\mathbf{U}}$. Let the eigenvalues of $\overline{\mathbf{R}}$ be denoted as $\bar{\lambda}_k = [\overline{\boldsymbol{\Lambda}}]_{k,k}$, and the diagonal entries of $\overline{\mathbf{Z}}$ as $\bar{z}_k = [\overline{\mathbf{Z}}]_{k,k}$, $k \in [0, M-1]$. Then the elements of $\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}$ can be written down in a more compact form:

$$
\begin{aligned}
[\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}]_{l_1,l_2} &= g_{l_1l_2} + \sum_{k=0}^{M-1}\bar{z}_k\left(\frac{g_{l_1}g_{l_2}(g_{l_1l_2}\bar{\lambda}_k+\rho_{l_1l_2})}{(g_{l_1}\bar{\lambda}_k+\rho_{l_1})(g_{l_2}\bar{\lambda}_k+\rho_{l_2})} - \frac{g_{l_1l_2}g_{l_1}}{g_{l_1}\bar{\lambda}_k+\rho_{l_1}} - \frac{g_{l_1l_2}g_{l_2}}{g_{l_2}\bar{\lambda}_k+\rho_{l_2}}\right) \\
&= g_{l_1l_2} - \sum_{k=0}^{M-1}\frac{\bar{z}_k[g_{l_1}g_{l_2}(g_{l_1l_2}\bar{\lambda}_k-\rho_{l_1l_2}) + g_{l_1l_2}(g_{l_1}\rho_{l_2}+g_{l_2}\rho_{l_1})]}{(g_{l_1}\bar{\lambda}_k+\rho_{l_1})(g_{l_2}\bar{\lambda}_k+\rho_{l_2})} \qquad . \qquad (3.101)
\end{aligned}
$$

3.4.4 MSE analysis of constrained MMSE-MMSE estimator

One can use the following brief description of the robust CMMSE-MMSE estimator that consists of the two successive processing modules and accepts the CLS estimate as an input:

$$
\hat{\underline{\mathbf{h}}}_m = \mathbf{W}\sum_{k=0}^{M-1}\overline{\boldsymbol{\Phi}}_k\,\tilde{\underline{\mathbf{h}}}_{m-k} \;, \qquad (3.102)
$$

where $\mathbf{W} = \mathbf{I} - \mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}(\mathbf{R}_{\underline{\mathbf{h}}\underline{\mathbf{h}}} + \mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}})^{-1}$, and the elements of $\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}$ are specified by (3.101), while $\overline{\boldsymbol{\Phi}}_k = \mathrm{dg}\left([\bar{f}_{0,k} \quad \cdots \quad \bar{f}_{L-1,k}]^{\mathrm{T}}\right)$ comprises FIR coefficients computed from $\bar{\mathbf{f}}_l = g_l(g_l\overline{\mathbf{R}}+\rho_l\mathbf{I})^{-1}\bar{\mathbf{r}}$ (3.67), with $\overline{\mathbf{R}} = (\mathrm{toe}[\bar{\mathbf{r}}, \bar{\mathbf{r}}^{\mathrm{H}}])^{\mathrm{T}}$ and the elements of $\bar{\mathbf{r}}$ being given by (3.90).

The channel estimation MSE can then be expressed as

$$
\begin{aligned}
\sigma_{\varepsilon}^2 &= \mathrm{tr}(\mathrm{E}[(\hat{\underline{\mathbf{h}}}_m - \underline{\mathbf{h}}_m)(\hat{\underline{\mathbf{h}}}_m - \underline{\mathbf{h}}_m)^{\mathrm{H}}]) \\
&= \mathrm{tr}(\mathrm{E}[\hat{\underline{\mathbf{h}}}_m\hat{\underline{\mathbf{h}}}_m^{\mathrm{H}}] + \mathrm{E}[\underline{\mathbf{h}}_m\underline{\mathbf{h}}_m^{\mathrm{H}}] - \mathrm{E}[\hat{\underline{\mathbf{h}}}_m\underline{\mathbf{h}}_m^{\mathrm{H}}] - \mathrm{E}[\underline{\mathbf{h}}_m\hat{\underline{\mathbf{h}}}_m^{\mathrm{H}}]) \\
&= \mathrm{tr}\left(\mathbf{W}\,\mathrm{E}\left[\left(\sum_{k_1=0}^{M-1}\overline{\boldsymbol{\Phi}}_{k_1}\tilde{\underline{\mathbf{h}}}_{m-k_1}\right)\left(\sum_{k_2=0}^{M-1}\tilde{\underline{\mathbf{h}}}_{m-k_2}^{\mathrm{H}}\overline{\boldsymbol{\Phi}}_{k_2}^{\mathrm{H}}\right)\right]\mathbf{W}^{\mathrm{H}} + \mathrm{E}[\underline{\mathbf{h}}_m\underline{\mathbf{h}}_m^{\mathrm{H}}] - \mathbf{W}\,\mathrm{E}\left[\sum_{k_1=0}^{M-1}\overline{\boldsymbol{\Phi}}_{k_1}\tilde{\underline{\mathbf{h}}}_{m-k_1}\underline{\mathbf{h}}_m^{\mathrm{H}}\right]\right) \\
&\quad - \mathrm{tr}\left(\mathrm{E}\left[\sum_{k_1=0}^{M-1}\underline{\mathbf{h}}_m\tilde{\underline{\mathbf{h}}}_{m-k_1}^{\mathrm{H}}\overline{\boldsymbol{\Phi}}_{k_1}^{\mathrm{H}}\right]\mathbf{W}^{\mathrm{H}}\right) \\
&= 1 + \mathrm{tr}\left[\mathbf{W}\left(\sum_{k_1=0}^{M-1}\sum_{k_2=0}^{M-1}\overline{\boldsymbol{\Phi}}_{k_1}\mathrm{E}[\tilde{\underline{\mathbf{h}}}_{m-k_1}\tilde{\underline{\mathbf{h}}}_{m-k_2}^{\mathrm{H}}]\overline{\boldsymbol{\Phi}}_{k_2}^{\mathrm{H}}\right)\mathbf{W}^{\mathrm{H}} - \mathbf{W}\sum_{k_1=0}^{M-1}\overline{\boldsymbol{\Phi}}_{k_1}\mathrm{E}[\underline{\mathbf{h}}_{m-k_1}\underline{\mathbf{h}}_m^{\mathrm{H}}] - \left(\mathbf{W}\sum_{k_1=0}^{M-1}\overline{\boldsymbol{\Phi}}_{k_1}\mathrm{E}[\underline{\mathbf{h}}_{m-k_1}\underline{\mathbf{h}}_m^{\mathrm{H}}]\right)^{\mathrm{H}}\right]. \qquad (3.103)
\end{aligned}
$$

Since the interblock CIR-plus-noise correlation matrix is equal to $\mathrm{E}[\underline{\tilde{\mathbf{h}}}_{m-k_1}\underline{\tilde{\mathbf{h}}}^{\mathrm{H}}_{m-k_2}] = r_{k_2-k_1}\mathbf{R}_{\underline{\mathbf{hh}}} + \delta_{k_2-k_1}SNR_{\mathrm{p}}^{-1}\mathbf{S}$, where $r_{k_2-k_1}$ is the true Doppler ACF evaluated in the interval $k_2-k_1 \in [-M+1, M-1]$, and $\delta_{k_2-k_1} = \begin{cases} 1, & k_2 = k_1 \\ 0, & k_2 \neq k_1 \end{cases}$ is the Dirac delta-function, it can be shown that

$$\sigma_{\varepsilon}^2 = 1 + \mathrm{tr}\left[\mathbf{W}\left(\sum_{k_1=0}^{M-1}\sum_{k_2=0}^{M-1} r_{k_2-k_1}\overline{\mathbf{\Phi}}_{k_1}\mathbf{R}_{\underline{\mathbf{hh}}}\overline{\mathbf{\Phi}}^{\mathrm{H}}_{k_2} + SNR_{\mathrm{p}}^{-1}\sum_{k=0}^{M-1}\overline{\mathbf{\Phi}}_k\mathbf{S}\overline{\mathbf{\Phi}}^{\mathrm{H}}_k\right)\mathbf{W}^{\mathrm{H}} - \mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{W}\sum_{k=0}^{M-1} r_{-k}\overline{\mathbf{\Phi}}_k - \left(\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{W}\sum_{k=0}^{M-1} r_{-k}\overline{\mathbf{\Phi}}_k\right)^{\mathrm{H}}\right]. \quad (3.104)$$

For an arbitrary pilot arrangement it is difficult to determine the upper and lower bounds on the estimation MSE (3.104). An exception is the comb pilot pattern, for which the upper bound is unambiguous and attained when $g_l = L^{-1}$ for $\forall l \in [0, L-1]$, as was established for both the CMMSE and CLS-MMSE estimators, which are the architectures underlying CMMSE-MMSE. Analogously to CMMSE and CLS-MMSE, the lower bound corresponds to the flat fading PDP. We leave derivation of the MSE bounds at the discretion of the reader.

3.4.5 Recursive operation mode

CLS-MMSE estimator (3.69) and CMMSE-MMSE estimator (3.98) rely on the a priori known second-order channel statistics, namely intrablock CIR correlation and SNR. The CLS-MMSE algorithm is much easier to implement in that sense as only PDP (not the entire CIR correlation matrix) is required for the Wiener filterbank design.

We have already suggested the efficient intrablock estimator architecture in Subsection 3.3.2, in case of the unknown channel statistics. Here we will show that a similar approach yields recursive two-dimensional estimation algorithms too.

Adopting the exponential averaging mechanism for the estimation of the CIR-plus-noise power profile at the filterbank input,

$$\hat{\tilde{g}}_{l,m} = (1-\alpha_{\mathrm{FF}})\,\hat{\tilde{g}}_{l,m-1} + \alpha_{\mathrm{FF}}\,|\underline{\tilde{h}}_{l,m}|^2, \quad (3.105)$$

we can redefine the coefficients of the *l*th robust filter (3.68) as

$$\hat{\tilde{\mathbf{f}}}_{l,m} = \hat{g}_{l,m}\sum_{k=0}^{M-1}\frac{\overline{w}_k}{(\hat{g}_{l,m}\overline{\lambda}_k + \rho_l)}\overline{\mathbf{u}}_k, \quad (3.106)$$

where $\overline{\lambda}_k$ and $\overline{\mathbf{u}}_k$ are the eigenvalues and eigenvectors of the flat-shaped Doppler correlation matrix $\overline{\mathbf{R}}$, $\overline{w}_k = \overline{\mathbf{u}}^{\mathrm{H}}_k\overline{\mathbf{r}}$,

$$\hat{g}_{l,m} = \begin{cases} \hat{\tilde{g}}_{l,m} - \rho_l, & \text{if } \hat{\tilde{g}}_{l,m} \geq \rho_l \\ 0, & \text{if } \hat{\tilde{g}}_{l,m} < \rho_l \end{cases} \quad (3.107)$$

is the PDP estimate obtained for the mth block, and the values of $\rho_l = \sigma_{\mathrm{w}}^2 \sigma_{\mathrm{p}}^{-2} [\mathbf{S}]_{l,l}$, $l \in [0, L-1]$ are specified for either a priori known WGN variance σ_{w}^2 or its estimate $\hat{\sigma}_{\mathrm{w}\,m}^2$, measured at the virtual subcarriers (refer to Subsection 3.3.2 for additional details). The forgetting factor α_{FF} can be adapted by the algorithm presented in Tab.3.1. Selection of $\hat{\tilde{g}}_{l,-1} = 0$ for $\forall l \in [0, L-1]$ is equivalent to the initial setting of the precision matrix in (3.26).

Based on the block-wise updated filterbank coefficients (3.106), one can define the *recursive CLS-MMSE* (RLS-MMSE) CFR estimate as

$$\hat{\mathbf{h}}_m^{\text{rls-mmse}} = \mathbf{F}\,\mathbf{B}\,\hat{\underline{\mathbf{h}}}_m^{\text{rls-mmse}} = \sigma_{\mathrm{p}}^{-2}\mathbf{F}\,\mathbf{B}\sum_{k=0}^{M-1}\hat{\tilde{\mathbf{\Phi}}}_{m,k}\mathbf{S}\,\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\,\mathrm{H}}\mathbf{y}^{\mathbf{p}}_{m-k}\,, \tag{3.108}$$

where $\hat{\tilde{\mathbf{\Phi}}}_{m,k} = \mathrm{dg}\left([\,\hat{\tilde{f}}_{0,m,k} \quad \cdots \quad \hat{\tilde{f}}_{L-1,m,k}\,]^{\mathrm{T}} \right)$. Implementation of the RLS-MMSE estimator is shown schematically in Fig.3.7.

Unlike the CLS-CMMSE estimator, where only PDP, $\mathrm{dg}(\mathbf{R}_{\underline{\mathbf{hh}}})$, needs to be acquired, the input error correlation matrix (3.101), contributing to design of the smoothing module in the CMMSE-MMSE implementation (3.98), depends on the off-diagonal elements of $\mathbf{R}_{\underline{\mathbf{hh}}}$. Hence one needs to compute the exponentially averaged estimate of the CIR-plus-noise correlation matrix

$$\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m} = (1-\alpha_{\mathrm{FF}})\,\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m-1} + \alpha_{\mathrm{FF}}\,\tilde{\underline{\mathbf{h}}}_m\,\tilde{\underline{\mathbf{h}}}_m^{\mathrm{H}}\,, \tag{3.109}$$

with $\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,-1} = \mathbf{0}_{L\times L}$, from where the estimated elements of the CIR correlation matrix $\mathbf{R}_{\underline{\mathbf{hh}}}$ are found as

$$\hat{g}_{l_1 l_2,\,m} = [\hat{\mathbf{R}}_{\underline{\mathbf{hh}}m}]_{l_1,l_2} = \begin{cases} [\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m}]_{l_1,l_2} - \rho_{l_1 l_2}\,, & \text{if } \min\limits_{l\in[0,L-1]} [\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m}]_{l,l} - \rho_l \geq 0 \\ [\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m}]_{l_1,l_2} - [\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m}]_{\tilde{l},\tilde{l}}\,[\mathbf{S}]^{-1}_{\tilde{l},\tilde{l}}\,[\mathbf{S}]_{l_1,l_2}\,, & \text{if } \min\limits_{l\in[0,L-1]} [\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m}]_{l,l} - \rho_l < 0 \end{cases}, \tag{3.110}$$

where $\tilde{l} = \arg\min\limits_{l\in[0,L-1]} [\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}\,m}]_{l,l} - \rho_l$. Conditioning in the expression (3.110) is necessary to prevent occurrence of the negative-definite estimates of the CIR correlation matrix $\hat{\mathbf{R}}_{\underline{\mathbf{hh}}m}$ that would contradict the theory. This is achieved by biasing the SNR value used in the estimator design to a lower setting (in analogy with the recommendation in Subsection 3.3.4.3).

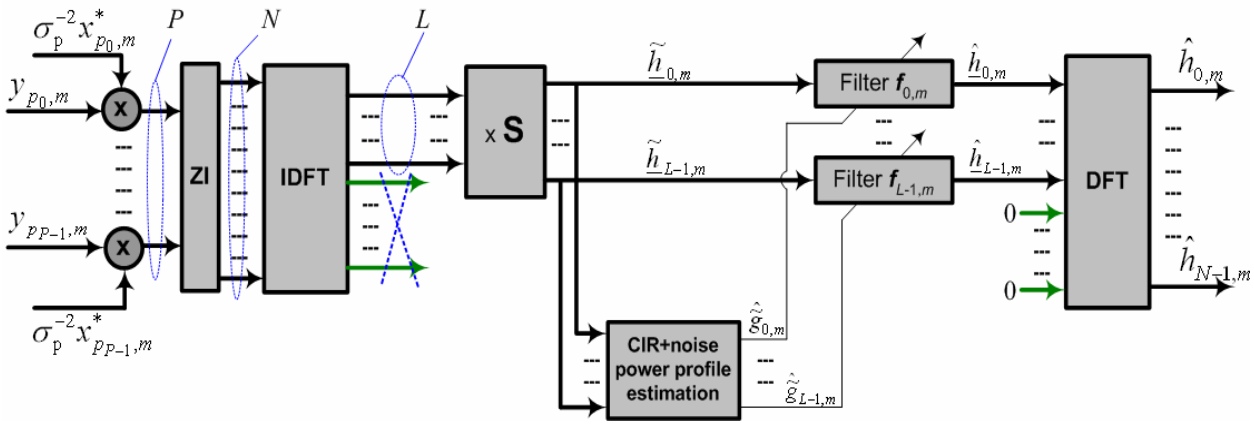


Fig.3.7. RLS-MMSE estimator

The estimates (3.110) are used in (3.101), so that

$$[\hat{\mathbf{R}}_{\underline{\varepsilon\varepsilon}\,m}]_{l_1,l_2} = \hat{g}_{l_1 l_2,m} - \sum_{k=0}^{M-1} \frac{\overline{z}_k [\hat{g}_{l_1 l_1,m}\hat{g}_{l_2 l_2,m}(\hat{g}_{l_1 l_2,m}\overline{\lambda}_k - \rho_{l_1 l_2}) + \hat{g}_{l_1 l_2,m}(\hat{g}_{l_1 l_1,m}\rho_{l_2 l_2} + \hat{g}_{l_2 l_2,m}\rho_{l_1 l_1})]}{(\hat{g}_{l_1 l_1,m}\overline{\lambda}_k + \rho_{l_1 l_1})(\hat{g}_{l_2 l_2,m}\overline{\lambda}_k + \rho_{l_2 l_2})}. \tag{3.111}$$

In analogy to the RMMSE implementation (Subsection 3.3.2), the matrix $\mathbf{T}_{\underline{\hat{h}}\underline{\hat{h}}} = (\mathbf{R}_{\underline{hh}} + \mathbf{R}_{\underline{\varepsilon\varepsilon}})^{-1}$ in (3.98) can be updated recursively block-by-block, yielding a reduced-complexity algorithm

$$\hat{\mathbf{T}}_{\underline{\hat{h}}\underline{\hat{h}}\,m} = \frac{1}{(1-\alpha_{\mathrm{FF}})}\hat{\mathbf{T}}_{\underline{\hat{h}}\underline{\hat{h}}\,m-1}\left[\mathbf{I} - \frac{\hat{\underline{\mathbf{h}}}_m \hat{\underline{\mathbf{h}}}_m^{\mathrm{H}}\hat{\mathbf{T}}_{\underline{\hat{h}}\underline{\hat{h}}\,m-1}}{(1-\alpha_{\mathrm{FF}})/\alpha_{\mathrm{FF}} + \mathrm{tr}(\hat{\underline{\mathbf{h}}}_m \hat{\underline{\mathbf{h}}}_m^{\mathrm{H}}\hat{\mathbf{T}}_{\underline{\hat{h}}\underline{\hat{h}}\,m-1})}\right], \tag{3.112}$$

where $\hat{\underline{\mathbf{h}}}_m = \hat{\underline{\mathbf{h}}}_m^{\text{rls-mmse}} = \sigma_{\mathrm{p}}^{-2}\sum_{k=0}^{M-1}\hat{\overline{\mathbf{\Phi}}}_{m,k}\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\,\mathrm{H}}\mathbf{y}^{\mathbf{p}}_{m-k}$, and $\hat{\mathbf{T}}_{\underline{\hat{h}}\underline{\hat{h}}-1} = \iota\mathbf{I}$ with $\iota \to \infty$.

Based on the design parameters specified in (3.106), (3.111) and (3.112), the *recursive CMMSE-MMSE* (RMMSE-MMSE) estimator is given as

$$\hat{\mathbf{h}}_m^{\text{rmmse-mmse}} = \mathbf{F}\mathbf{B}\hat{\underline{\mathbf{h}}}_m^{\text{rmmse-mmse}} = \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}(\mathbf{I} - \hat{\mathbf{R}}_{\underline{\varepsilon\varepsilon}\,m}\hat{\mathbf{T}}_{\underline{\hat{h}}\underline{\hat{h}}\,m})\sum_{k=0}^{M-1}\hat{\overline{\mathbf{\Phi}}}_{m,k}\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\,\mathrm{H}}\mathbf{y}^{\mathbf{p}}_{m-k}, \tag{3.113}$$

and is shown in the form of the functional diagram in Fig.3.8. It can be seen clearly that the estimator consists of the three successive modules: CLS weighting, Wiener filterbank and intrablock smoothing.

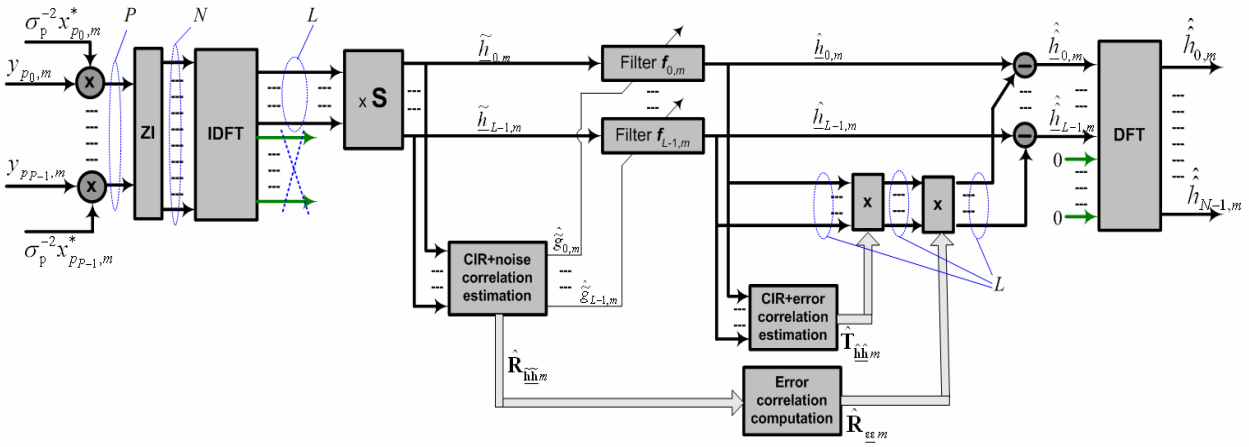


Fig.3.8. RMMSE-MMSE estimator

3.4.6 Computational complexity

The input of the Wiener filterbank in the CLS-MMSE estimator (3.69) and CMMSE-MMSE estimator (3.98), as well as of their recursive counterparts, (3.108) and (3.113), represents the CLS estimate of CIR, requiring in general L^2 multiplications and $L^2 - L$ additions for the mth block, in addition to the IFFT processing effort, as described in Subsections 3.2.2 and 3.3.3. CLS estimates of CIR, corresponding to $M-1$ preceding blocks, also used in filtering, are stored in the buffer. In the special case of the comb pilot pattern, no computations are needed to produce the CLS estimates at the IFFT output.

In case of the fixed precalculated set of FIR coefficients for each of the Wiener filters, the corresponding stage of the CLS-MMSE estimator involves only $M\,L$ multiplies and $(M-1)\,L$ additions. For the precomputed CMMSE-MMSE estimator, the multiplicative and additive complexities of the Wiener filtering and the smoothing module are $M\,L + L^2$ and $(M-2)\,L + L^2$, respectively.

CLS-MMSE and CMMSE-MMSE estimators, relying on the regular channel statistics update, are more complicated, due to the necessity to calculate M FIR coefficients for each of L filters in the filterbank. According to (3.68), calculation of a single coefficient $f_{l,k}$, $k \in [0, M-1]$ incurs $2M+1$ multiplies and M additions. This can make implementation of the estimator challenging if the desired filter length M is large. A deeper investigation reveals that magnitudes of the eigenvalues $\lambda_k = \overline{\lambda}_k$ of the Doppler correlation matrix $\mathbf{R} = \overline{\mathbf{R}}$ in (3.68) rapidly descend to zero if the maximum Doppler shift $\overline{\omega}_{\mathrm{D}}$, used for the robust filterbank design in (3.89)-(3.90), is relatively small. Thus, the set of the robust Wiener FIR coefficients (3.89) and their estimates (3.106) can be computed correspondingly as

$$\bar{\mathbf{f}}_l \approx \frac{g_l}{\rho_l}\left(\sum_{k=0}^{M'-1}\frac{\overline{w}_k}{(\overline{\lambda}_k\ g_l/\rho_l+1)}\overline{\mathbf{u}}_k+\sum_{k=M'}^{M-1}\overline{w}_k\overline{\mathbf{u}}_k\right), \tag{3.114}$$

$$\hat{\bar{\mathbf{f}}}_{l,m} \approx \frac{\hat{g}_{l,m}}{\rho_l}\left(\sum_{k=0}^{M'-1}\frac{\overline{w}_k}{(\overline{\lambda}_k\ \hat{g}_{l,m}/\rho_l+1)}\overline{\mathbf{u}}_k+\sum_{k=M'}^{M-1}\overline{w}_k\overline{\mathbf{u}}_k\right), \tag{3.115}$$

where the approximation of $\overline{\lambda}_k \approx 0$, $k \in [M', M-1]$ does not cause performance degradation if M' is properly selected. Fig.3.9 shows the minimum values of M', which entail eigenvalue magnitude decrease below 0.001% of the maximum eigenvalue, determined for the range of $\overline{\omega}_{\mathrm{D}} \in [0.025\pi, 0.5\pi]$ and $M \in [10, 265]$. One can see that the rank of the Doppler correlation matrix can be approximated as

$$M' = \lceil M\,\overline{\omega}_{\mathrm{D}}/\pi \rceil + \Delta, \tag{3.116}$$

where Δ value is selected from 3 to 9 according to Fig.3.9.

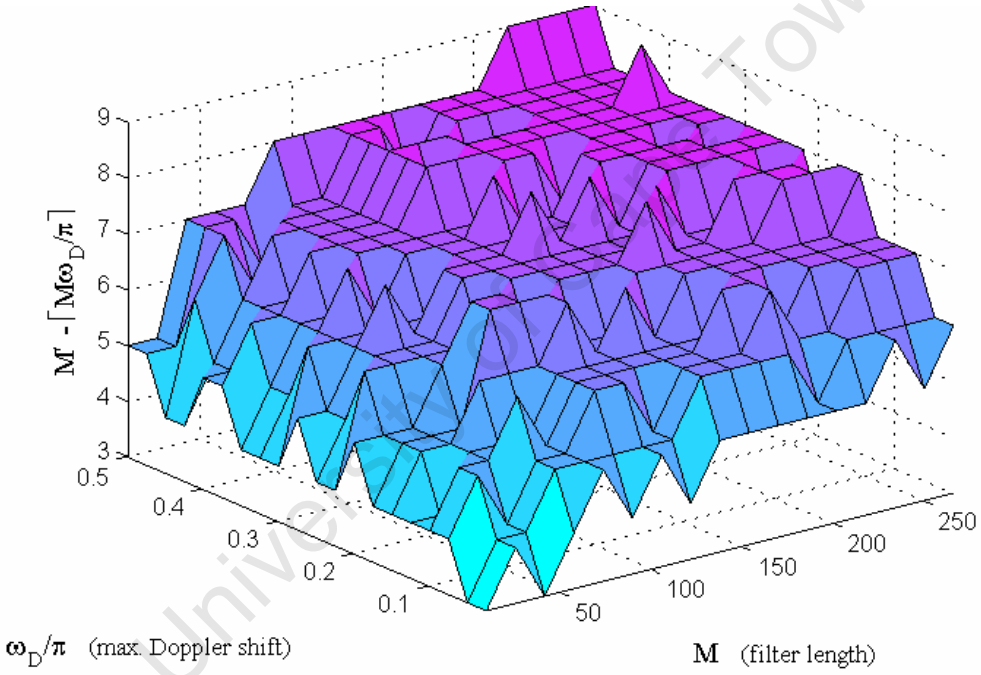


Fig.3.9. Approximated rank of the Doppler correlation matrix used for the robust filterbank design, determined for the eigenvalue magnitude fall below 0.001% of the maximum (here $\omega_{\mathrm{D}} = \overline{\omega}_{\mathrm{D}}$)

Assuming that PDP ($g_l, l \in [0, L-1]$) and SNR ($\rho_l, l \in [0, L-1]$) in (3.114) are acquired block-wise using external means, we can express the number of multiples to obtain $\bar{\mathbf{f}}_l$ as $M' + M'M + M + 1 = M(M'+1) + M' + 1$ and the number of additions as $M' + M'M$. The RLS-MMSE implementation includes g_l estimation by means of exponential averaging (3.105) that represents 3 extra multiplications and 2 extra additions per filterbank branch. The total number of mathematical operations for the CLS-MMSE and RLS-MMSE estimator is shown in Tab.3.3.

CMMSE-MMSE estimator shares the same complexity as CLS-MMSE for the filterbank module plus the computational demand of the smoothing module, which is similar to the CMMSE architecture. Recalling the rank reduction property (3.116) of the Doppler correlation matrix, one can compute the elements of the error correlation matrix $\mathbf{R}_{\underline{\varepsilon\varepsilon}}$ (3.101) and their estimates (3.111) as

$$[\mathbf{R}_{\underline{\varepsilon\varepsilon}}]_{l_1,l_2} \approx g_{l_1l_2} - \sum_{k=0}^{M'-1} \frac{\overline{z}_k [g_{l_1} g_{l_2} g_{l_1l_2} \overline{\lambda}_k + g_{l_1l_2}(g_{l_1}\rho_{l_2} + g_{l_2}\rho_{l_1}) - g_{l_1} g_{l_2} \rho_{l_1l_2}]}{(g_{l_1}\overline{\lambda}_k + \rho_{l_1})(g_{l_2}\overline{\lambda}_k + \rho_{l_2})}$$
$$- \frac{g_{l_1l_2}(g_{l_1}\rho_{l_2} + g_{l_2}\rho_{l_1}) - g_{l_1} g_{l_2} \rho_{l_1l_2}}{\rho_{l_1}\rho_{l_2}} \sum_{k=M'}^{M-1} \overline{z}_k \quad , \tag{3.117}$$

$$[\hat{\mathbf{R}}_{\underline{\varepsilon\varepsilon}m}]_{l_1,l_2} \approx \hat{g}_{l_1l_2,m} - \sum_{k=0}^{M'-1} \frac{\overline{z}_k [\hat{g}_{l_1l_1,m}\hat{g}_{l_2l_2,m}(\hat{g}_{l_1l_2,m}\overline{\lambda}_k - \rho_{l_1l_2}) + \hat{g}_{l_1l_2,m}(\hat{g}_{l_1l_1,m}\rho_{l_2l_2} + \hat{g}_{l_2l_2,m}\rho_{l_1l_1})]}{(\hat{g}_{l_1l_1,m}\overline{\lambda}_k + \rho_{l_1l_1})(\hat{g}_{l_2l_2,m}\overline{\lambda}_k + \rho_{l_2l_2})}$$
$$- \frac{\hat{g}_{l_1l_2,m}(\hat{g}_{l_1l_1,m}\rho_{l_2l_2} + \hat{g}_{l_2l_2,m}\rho_{l_1l_1}) - \hat{g}_{l_1l_1,m}\hat{g}_{l_2l_2,m}\rho_{l_1l_2}}{\rho_{l_1l_1}\rho_{l_2l_2}} \sum_{k=M'}^{M-1} \overline{z}_k \quad . \tag{3.118}$$

We first consider the estimator with external acquisition of $\mathbf{R}_{\underline{\mathbf{hh}}}$ and $SNR_{\mathrm{p}}^{-1}\mathbf{S}$. Each element of the Hermitean matrix $\mathbf{R}_{\underline{\varepsilon\varepsilon}}$ (3.117) requires $6M'+8$ multiplications and $3M'+4$ additions. Thus, the overall complexity of $\mathbf{R}_{\underline{\varepsilon\varepsilon}}$ computation is determined by processing $\sum_{l=1}^{L} l = L(L+1)/2$ elements, and constitutes $(3M'+4)(L^2+L)$ multiplies and $(1.5M'+2)(L^2+L)$ additions. In analogy with the CMMSE estimator (refer to Subsection 3.3.3), the inverted weighting matrix $(\mathbf{R}_{\underline{\mathbf{hh}}} + \mathbf{R}_{\underline{\varepsilon\varepsilon}})^{-1}$ can be found using the Gauss-Jordan elimination (Appendix B.1), which requires L^3 multiplies and $L^3 - 2L^2 + L$ additions. Hence the product $[\mathbf{I} - \mathbf{R}_{\underline{\varepsilon\varepsilon}}(\mathbf{R}_{\underline{\mathbf{hh}}} + \mathbf{R}_{\underline{\varepsilon\varepsilon}})^{-1}]\hat{\underline{\mathbf{h}}}_m^{\text{cls-mmse}}$ (3.98) needs $L^3 + 2L^2$ multiplies and $0.5(L^2+L) + L^3 - 2L^2 + L + 2(L^2-L) + L = L^3 + 0.5L^2 + 0.5L$ additions. The obtained figures should be added to the Wiener filterbank complexity to get the expressions of the total number of mathematical operations required by the estimator.

RMMSE-MMSE implementation involves two exponential averaging procedures. The first one needs $3(L^2+L)/2$ multiplications and $(L^2+L)/2$ additions to compute $\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}m}$ (3.109), plus $(L^2+L)/2$ additions to obtain $\hat{\mathbf{R}}_{\underline{\mathbf{hh}}m}$, the elements of which are used in computation of $\hat{\mathbf{R}}_{\underline{\varepsilon\varepsilon}m}$ (3.118). The second averaging underlies a block-wise update of the precision matrix $\mathbf{T}_{\underline{\hat{\mathbf{h}}\hat{\mathbf{h}}}} = (\mathbf{R}_{\underline{\mathbf{hh}}} + \mathbf{R}_{\underline{\varepsilon\varepsilon}})^{-1}$ in the smoothing module described by (3.112). In analogy with the RMMSE algorithm (refer to Subsection 3.3.3), $5L^2 + 2L$ multiplies and $4L^2 - L$ additions are needed to compute $[\mathbf{I} - (1-\alpha_{\mathrm{FF}})^{-1}\hat{\mathbf{R}}_{\underline{\varepsilon\varepsilon}m}\hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}\hat{\mathbf{h}}}m-1}\mathbf{W}_m]\hat{\underline{\mathbf{h}}}_m$, where $\mathbf{W}_m = \mathbf{I} - \frac{\hat{\underline{\mathbf{h}}}_m \hat{\underline{\mathbf{h}}}_m^{\mathrm{H}} \hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}\hat{\mathbf{h}}}m-1}}{(1-\alpha_{\mathrm{FF}})/\alpha_{\mathrm{FF}} + \mathrm{tr}(\hat{\underline{\mathbf{h}}}_m \hat{\underline{\mathbf{h}}}_m^{\mathrm{H}} \hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}\hat{\mathbf{h}}}m-1})}$. In the remaining aspects, RMMSE-MMSE complexity is identical to that of the CMMSE-MMSE algorithm. The summary of the estimator complexities is presented in Tab.3.3.

Tab.3.3. Computational complexity for the 2D estimator family

Estimator	Multiplications (C_{M})	Additions (C_{A})
CLS-MMSE (fixed $\bar{\mathbf{f}}_l$)	$C_{\mathrm{M}}^{\mathrm{FFT}} + M\,L + L^2$	$C_{\mathrm{A}}^{\mathrm{FFT}} + (M-2)\,L + L^2$
CLS-MMSE (updated $\bar{\mathbf{f}}_l$)	$C_{\mathrm{M}}^{\mathrm{FFT}} + (M'+2)\,M\,L + (M'+1)\,L + L^2$	$C_{\mathrm{A}}^{\mathrm{FFT}} + (M'+1)\,M\,L + (M'-2)\,L + L^2$
RLS-MMSE	$C_{\mathrm{M}}^{\mathrm{FFT}} + (M'+2)\,M\,L + (M'+4)\,L + L^2$	$C_{\mathrm{A}}^{\mathrm{FFT}} + (M'+1)\,M\,L + M'\,L + L^2$
CMMSE-MMSE (fixed $\bar{\mathbf{f}}_l$ and $[\mathbf{I}-\mathbf{R}_{\underline{\varepsilon\varepsilon}}(\mathbf{R}_{\underline{\mathbf{hh}}}+\mathbf{R}_{\underline{\varepsilon\varepsilon}})^{-1}]$)	$C_{\mathrm{M}}^{\mathrm{FFT}} + M\,L + 2L^2$	$C_{\mathrm{A}}^{\mathrm{FFT}} + (M-3)\,L + 2L^2$
CMMSE-MMSE (upd. $\bar{\mathbf{f}}_l$ and $[\mathbf{I}-\mathbf{R}_{\underline{\varepsilon\varepsilon}}(\mathbf{R}_{\underline{\mathbf{hh}}}+\mathbf{R}_{\underline{\varepsilon\varepsilon}})^{-1}]$)	$C_{\mathrm{M}}^{\mathrm{FFT}} + (M'+2)\,M\,L + (3M'+7)\,L^2 + (4M'+5)\,L + L^3$	$C_{\mathrm{A}}^{\mathrm{FFT}} + (M'+1)\,M\,L + (1.5M'+3.5)\,L^2 + (2.5M'+0.5)\,L + L^3$
RMMSE-MMSE	$C_{\mathrm{M}}^{\mathrm{FFT}} + (M'+2)\,M\,L + (3M'+11.5)\,L^2 + (4M'+11.5)\,L$	$C_{\mathrm{A}}^{\mathrm{FFT}} + (M'+1)\,M\,L + (1.5M'+8)\,L^2 + (2.5M'+2)\,L$

3.4.7 Extension to MIMO

Analysing the error correlation of CLS estimates, forming the input of the filtering module of CLS-MMSE (refer to Subsection 3.2.3), one should note that the doubly symmetric matrices $\mathbf{S}(i)$ for all $i \in [0, N_{\mathrm{tx}}-1]$ have the same eigenvalues and the same elements on the main diagonal, though the off-diagonal elements are different in general. Equality of the corresponding diagonal elements, $\rho_l(i) = SNR_{\mathrm{p}}^{-1}[\mathbf{S}(i)]_{l,l} = SNR_{\mathrm{p}}^{-1}[\mathbf{S}(0)]_{l,l} = \rho_l$ for $\forall i \in [0, N_{\mathrm{tx}}-1]$, implies Wiener filterbanks being identical for all SLs. Hence the vector of coefficients of the *l*th filterbank branch for the (i, j) th SL is simply expressed as (3.89):

$$\bar{\mathbf{f}}_l(i) = \bar{\mathbf{f}}_l = g_l(g_l\overline{\mathbf{R}} + \rho_l\mathbf{I})^{-1}\bar{\mathbf{r}} = g_l\sum_{k=0}^{M-1}\frac{\overline{w}_k}{(g_l\overline{\lambda}_k + \rho_l)}\overline{\mathbf{u}}_k\ , \tag{3.119}$$

where $\overline{\lambda}_k$ and $\overline{\mathbf{u}}_k$ are the eigenvalues and eigenvectors of the flat-shape Doppler correlation matrix $\overline{\mathbf{R}}$, and $\overline{w}_k = \overline{\mathbf{u}}_k^{\mathrm{H}}\bar{\mathbf{r}}$. The aforementioned property makes the CLS-MMSE implementation attractive from the low complexity standpoint, irrespective of the pilot pattern adopted. The CLS-MMSE CFR estimate for the (i, j) th SL is expressed as

$$\hat{\mathbf{h}}_m^{\text{cls-mmse}}(j,i) = \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}\sum_{k=0}^{M-1}\overline{\mathbf{\Phi}}_{m,k}\mathbf{S}(i)\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}(i)^{\mathrm{H}}\mathbf{X}^{\mathbf{p}}(i)^{\mathrm{H}}\mathbf{y}_{m-k}^{\mathbf{p}}(j,i)\ , \tag{3.120}$$

where $\overline{\mathbf{\Phi}}_{m,k} = \mathrm{dg}\left([\bar{f}_{0,m,k} \quad \cdots \quad \bar{f}_{L-1,m,k}]^{\mathrm{T}}\right)$.

In contrast to CLS-MMSE, design of the smoothing module of the CMMSE-MMSE estimator is SL-specific, with the underlying error correlation matrix at the input of the smoothing module for the (i, j) th SL being given as

$$[\mathbf{R}_{\underline{\varepsilon\varepsilon}}(i)]_{l_1,l_2} = g_{l_1l_2} - \sum_{k=0}^{M-1}\frac{\overline{z}_k[g_{l_1}g_{l_2}(g_{l_1l_2}\overline{\lambda}_k - \rho_{l_1l_2}(i)) + g_{l_1l_2}(g_{l_1}\rho_{l_2} + g_{l_2}\rho_{l_1})]}{(g_{l_1}\overline{\lambda}_k + \rho_{l_1})(g_{l_2}\overline{\lambda}_k + \rho_{l_2})}\ , \tag{3.121}$$

where $\rho_{l_1 l_2}(i) = SNR_{\mathrm{p}}^{-1}[\mathbf{S}(i)]_{l_1,l_2}$ (3.93). The CMMSE-MMSE CFR estimate for the (i, j) th SL can be described as

$$\hat{\mathbf{h}}_m^{\text{cmmse-mmse}}(j,i) = \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}[\mathbf{I} - \mathbf{R}_{\underline{\varepsilon\varepsilon}}(i)(\mathbf{R}_{\underline{\mathbf{hh}}} + \mathbf{R}_{\underline{\varepsilon\varepsilon}}(i))^{-1}]\sum_{k=0}^{M-1}\overline{\mathbf{\Phi}}_{m,k}\mathbf{S}(i)\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}(i)^{\mathrm{H}}\mathbf{X}^{\mathbf{p}}(i)^{\mathrm{H}}\mathbf{y}_{m-k}^{\mathbf{p}}(j,i)\,. \tag{3.122}$$

In analogy with the RMMSE estimator (refer to Subsection 3.3.5), in the MIMO RLS-MMSE implementation, coefficients of the Wiener filterbank can be updated block-wise, taking into account channel statistics collection from all SLs. Namely the exponentially averaged CIR-plus-noise power profile (3.105) benefits from additional smoothing over $N_{\mathrm{tx}}N_{\mathrm{rx}}$ independent layers:

$$\hat{\bar{g}}_{l,m} = (1-\alpha_{\mathrm{FF}})\,\hat{\bar{g}}_{l,m-1} + \frac{\alpha_{\mathrm{FF}}}{N_{\mathrm{rx}}N_{\mathrm{tx}}}\sum_{i=0}^{N_{\mathrm{tx}}-1}\sum_{j=0}^{N_{\mathrm{rx}}-1}|\,\underline{\tilde{h}}_{l,m}(j,i)\,|^2\,, \tag{3.123}$$

where $\underline{\tilde{h}}_{l,m}(j,i)$ is the CLS estimate of the lth CIR sample, obtained for the (i, j) th SL. Based on the PDP estimate

$$\hat{g}_{l,m} = \begin{cases} \hat{\bar{g}}_{l,m} - \rho_l, & \text{if } \hat{\bar{g}}_{l,m} \geq \rho_l \\ 0, & \text{if } \hat{\bar{g}}_{l,m} < \rho_l \end{cases}$$ (3.107), we can express the RLS-MMSE estimate of CFR:

$$\hat{\mathbf{h}}_m^{\text{rls-mmse}}(j,i) = \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}\sum_{k=0}^{M-1}\hat{\overline{\mathbf{\Phi}}}_{m,k}\mathbf{S}(i)\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}(i)^{\mathrm{H}}\mathbf{X}^{\mathbf{p}}(i)^{\mathrm{H}}\mathbf{y}_{m-k}^{\mathbf{p}}(j,i)\,, \tag{3.124}$$

where $\hat{\overline{\mathbf{\Phi}}}_{m,k} = \mathrm{dg}\left([\,\hat{\bar{f}}_{0,m,k} \quad \cdots \quad \hat{\bar{f}}_{L-1,m,k}\,]^{\mathrm{T}}\right)$, and $\hat{\bar{\mathbf{f}}}_{l,m} = \hat{g}_{l,m}\sum_{k=0}^{M-1}\frac{\overline{w}_k}{(\hat{g}_{l,m}\overline{\lambda}_k + \rho_l)}\overline{\mathbf{u}}_k$.

The recursive implementation of the smoothing module, being a part of the RMMSE-MMSE estimator, can be described as follows, similar to (3.105) and (3.53)-(3.54):

$$\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(i) = (1-\alpha_{\mathrm{FF}})\,\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m-1}(i) + \frac{\alpha_{\mathrm{FF}}}{N_{\mathrm{rx}}}\sum_{j=0}^{N_{\mathrm{rx}}-1}\underline{\tilde{\mathbf{h}}}_m(j,i)\underline{\tilde{\mathbf{h}}}_m(j,i)^{\mathrm{H}}\,, \tag{3.125}$$

$$\hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}}\underline{\hat{\mathbf{h}}}m}(i) = \frac{1}{(1-\alpha_{\mathrm{FF}})}\hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}}\underline{\hat{\mathbf{h}}}m-1}(i)\left[\mathbf{I} - \frac{\mathbf{V}_m(i)}{N_{\mathrm{rx}}(1-\alpha_{\mathrm{FF}})/\alpha_{\mathrm{FF}} + \mathrm{tr}[\mathbf{V}_m(i)]}\right], \tag{3.126}$$

where $\mathbf{V}_m(i) = \left(\sum_{j=0}^{N_{\mathrm{rx}}-1}\underline{\hat{\mathbf{h}}}_m(j,i)\underline{\hat{\mathbf{h}}}_m(j,i)^{\mathrm{H}}\right)\hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}}\underline{\hat{\mathbf{h}}}m-1}(i)$. Based on the parameters $\hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}}\underline{\hat{\mathbf{h}}}m}(i)$ and

$$\hat{\mathbf{R}}_{\underline{\mathbf{hh}}m} = \begin{cases} N_{\mathrm{tx}}^{-1}\sum_{i=0}^{N_{\mathrm{tx}}-1}\left(\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(i) - SNR_{\mathrm{p}}^{-1}\mathbf{S}(i)\right), & \text{if } \min\limits_{l\in[0,L-1]}\sum_{i=0}^{N_{\mathrm{tx}}-1}\left([\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(i)]_{l,l} - SNR_{\mathrm{p}}^{-1}[\mathbf{S}(i)]_{l,l}\right) \geq 0 \\ N_{\mathrm{tx}}^{-1}\sum_{i=0}^{N_{\mathrm{tx}}-1}\left(\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(i) - [\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(i)]_{\tilde{l},\tilde{l}}[\mathbf{S}(i)]_{\tilde{l},\tilde{l}}^{-1}\mathbf{S}(i)\right), & \text{if } \min\limits_{l\in[0,L-1]}\sum_{i=0}^{N_{\mathrm{tx}}-1}\left([\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(i)]_{l,l} - SNR_{\mathrm{p}}^{-1}[\mathbf{S}(i)]_{l,l}\right) < 0 \end{cases}$$, where

$\tilde{l} = \arg\min\limits_{l\in[0,L-1]}\sum_{i=0}^{N_{\mathrm{tx}}-1}\left([\hat{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}m}(i)]_{l,l} - SNR_{\mathrm{p}}^{-1}[\mathbf{S}(i)]_{l,l}\right)$ (3.110), the RMMSE-MMSE CFR estimate for the (i, j) th SL is described as

$$\hat{\mathbf{h}}_m^{\text{rmmse-mmse}}(j,i) = \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}[\mathbf{I} - \hat{\mathbf{R}}_{\underline{\varepsilon\varepsilon}m}(i)\,\hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}}\underline{\hat{\mathbf{h}}}m}(i)]\sum_{k=0}^{M-1}\hat{\overline{\mathbf{\Phi}}}_{m,k}\mathbf{S}(i)\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}(i)^{\mathrm{H}}\mathbf{X}^{\mathbf{p}}(i)^{\mathrm{H}}\mathbf{y}_{m-k}^{\mathbf{p}}(j,i)\,, \tag{3.127}$$

where

$$[\hat{\mathbf{R}}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}m}(i)]_{l_1,l_2} = \hat{g}_{l_1l_2,m} - \sum_{k=0}^{M-1} \frac{\overline{z}_k [\hat{g}_{l_1l_1,m}\hat{g}_{l_2l_2,m}(\hat{g}_{l_1l_2,m}\overline{\lambda}_k - \rho_{l_1l_2}(i)) + \hat{g}_{l_1l_2,m}(\hat{g}_{l_1l_1,m}\rho_{l_2l_2} + \hat{g}_{l_2l_2,m}\rho_{l_1l_1})]}{(\hat{g}_{l_1l_1,m}\overline{\lambda}_k + \rho_{l_1l_1})(\hat{g}_{l_2l_2,m}\overline{\lambda}_k + \rho_{l_2l_2})}, \tag{3.128}$$

and $\hat{g}_{l_1l_2,m} = [\hat{\mathbf{R}}_{\underline{\mathbf{h}}\underline{\mathbf{h}}m}]_{l_1,l_2}$.

As shown in Appendix C.2, the equispaced pilot pattern, implying equality of the weighting matrices $\mathbf{S}(i) = \mathbf{S}(0) = P^{-1}\mathbf{I}$ for $\forall i \in [0, N_{\text{tx}} - 1]$, lowers receiver complexity due to $\mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}(i) = \mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}(0) = \mathbf{R}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}}$, $\hat{\mathbf{R}}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}m}(i) = \hat{\mathbf{R}}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}m}(0) = \hat{\mathbf{R}}_{\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}m}$ and $\hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}}\underline{\hat{\mathbf{h}}}m}(i) = \hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}}\underline{\hat{\mathbf{h}}}m}(0) = \hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}}\underline{\hat{\mathbf{h}}}m}$ in (3.121), (3.128) and (3.126). An improved convergence of the RMMSE-MMSE estimator is achieved by additional averaging of the precision matrix estimates $\hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}}\underline{\hat{\mathbf{h}}}m}$ across SLs corresponding to different Tx antennas.

3.5 Comparison of estimator complexities

The best channel estimator cannot be selected solely based on the MSE performance. For instance, it has already been mentioned in Subsection 3.4.2 that an ideal interblock estimator cannot be implemented in hardware due to infinite memory requirements. Furthermore, the bulky filter architectures are not desirable for the high-dimensional MIMO transceivers. On the other hand, increase of the processing block size in prospective systems (e.g., UWB multicarrier architectures) potentially broadens the parameter space considered in the intrablock CIR estimation. The latter circumstance poses a question as to the reasonable complexity level for the intrablock estimators, and whether the interblock extension can improve estimation accuracy without the considerable growth of computation load due to the increased number of filters in the filterbank.

It might seem at first sight that complexity of the 2D estimators is always higher than that of the intrablock estimators. This statement is indeed true for the CMMSE-MMSE and RMMSE-MMSE algorithms, which incorporate the CMMSE and RMMSE modules being known as the most computationally demanding intrablock estimators (refer to Tab.3.2). However, comparing between CMMSE/RMMSE and CLS-MMSE, there could be different variants, depending on the anticipated CIR length L, the filterbank length M, and the effective rank of the Doppler correlation matrix M'.

Tab.3.4 compares the most complexity-effective fully functional implementations of the 1D and 2D estimators described in Sections 3.2, 3.3 and 3.4 (refer to Tab.3.2 and Tab.3.3). It can be seen that the order of complexity with regard to the CIR length does not exceed $O(L^2)$. For the 2D estimators, the order of complexity is defined as $O(M'M\,L) + O(M'L^2)$. Thus it matters what is bigger – M or L, – to establish the dominating complexity factor. One can also see that the CLS estimator serves as the lowest complexity reference, whereas the RMMSE-MMSE estimator tops the list and is supposed to be used in the receivers aiming at ultimate performance.

Tab.3.4. Comparative summary of estimator complexities

Estimator	Multiplications (C_{M})	Additions (C_{A})
CLS	$C_{\mathrm{M}}^{\mathrm{FFT}} + L^2$	$C_{\mathrm{A}}^{\mathrm{FFT}} + L^2 - L$
RMMSE	$C_{\mathrm{M}}^{\mathrm{FFT}} + 6L^2 + 3L$	$C_{\mathrm{A}}^{\mathrm{FFT}} + 5L^2 - 2L$
RLS-MMSE	$C_{\mathrm{M}}^{\mathrm{FFT}} + (M'+2)ML + (M'+4)L + L^2$	$C_{\mathrm{A}}^{\mathrm{FFT}} + (M'+1)ML + M'L + L^2$
RMMSE-MMSE	$C_{\mathrm{M}}^{\mathrm{FFT}} + (M'+2)ML + (3M'+11.5)L^2 + (4M'+11.5)L$	$C_{\mathrm{A}}^{\mathrm{FFT}} + (M'+1)ML + (1.5M'+8)L^2 + (2.5M'+2)L$

For better understanding of the impact of the parameters L, M and M' on the overall complexity, consider an example of the system with $N = 64$ subcarriers and $P \geq 16$ arbitrarily spaced pilots that is designed for $\overline{\omega}_{\mathrm{D}} = 0.025\pi$. We can analyse several typical CIR length settings:

1) if $L = 8$, then the CLS estimator requires $C_{\mathrm{M}}^{\mathrm{FFT}} + 64$ multiplications and $C_{\mathrm{A}}^{\mathrm{FFT}} + 56$ additions, RMMSE is characterised by $C_{\mathrm{M}}^{\mathrm{FFT}} + 408$ multiplies and $C_{\mathrm{A}}^{\mathrm{FFT}} + 304$ additions, the RLS-MMSE estimator equivalent to RMMSE should have $M = 7$, for which $M' = 3$ (3.116), resulting in $C_{\mathrm{M}}^{\mathrm{FFT}} + 400$ multiplies and $C_{\mathrm{A}}^{\mathrm{FFT}} + 312$ additions, and the RMMSE-MMSE estimator with the same filterbank ($M = 7$) requires $C_{\mathrm{M}}^{\mathrm{FFT}} + 1780$ multiplies and $C_{\mathrm{A}}^{\mathrm{FFT}} + 1100$ additions;
2) if $L = 16$, then the CLS estimator requires $C_{\mathrm{M}}^{\mathrm{FFT}} + 256$ multiplications and $C_{\mathrm{A}}^{\mathrm{FFT}} + 240$ additions, RMMSE is characterised by $C_{\mathrm{M}}^{\mathrm{FFT}} + 1584$ multiplies and $C_{\mathrm{A}}^{\mathrm{FFT}} + 1248$ additions, the RLS-MMSE estimator equivalent to RMMSE should have $M = 12$, for which $M' = 4$, resulting in $C_{\mathrm{M}}^{\mathrm{FFT}} + 1536$ multiplies and $C_{\mathrm{A}}^{\mathrm{FFT}} + 1280$ additions, and the RMMSE-MMSE estimator with the same filterbank ($M = 12$) requires $C_{\mathrm{M}}^{\mathrm{FFT}} + 7608$ multiplies and $C_{\mathrm{A}}^{\mathrm{FFT}} + 4736$ additions.

One can see that the RMMSE complexity increases rapidly with L, and the computationally equivalent RLS-MMSE estimator can be implemented with longer memory (M) that makes it a reasonable alternative to RMMSE for bigger L values.

3.6 Numerical MSE performance

Randomness and variety of the wireless channels poses a question as to what channel estimator is applicable to what scenario. An essential comparison criterion is the MSE performance. It has been shown throughout this chapter that the analytical MSE expressions serve as a convenient tool for evaluation of the optimal and suboptimal pilot-assisted estimators, which rely on the known channel statistics (e.g., CMMSE, CLS-MMSE, etc.). A detailed

comparative study of the corresponding estimation algorithms is presented in Subsection 3.6.2. We emphasise the effect of the multipath order (the number of independent multipath components) on the estimator performance and discuss appropriate receiver architectures, taking into account design complexity. Although this thesis does not include the analytical MSE expressions for the recursive versions of the estimation algorithms, i.e. those, which rely on the channel statistics estimation over the block sequence, their MSE performance is investigated by means of the simulation experiments in Subsection 3.6.3. The main objective here is to show that the steady-state performance of the recursive algorithms approaches that of their theoretical counterparts, designed based on the known channel statistics. Finally, it should be noted that the symbol error rate (SER) performance is not of interest in the context of the current chapter. This will be presented in Chapter 4, which is dedicated to the overall system optimisation.

3.6.1 System configuration

Similar to the experiments in Section 2.4, we consider a discrete-time fully synchronised baseband OFDM system with the processing block length $N = 64$. However, in the following analysis the CP length is set to $N_{\text{cp}} = 15$ to accommodate CIR with the modelled length of $L = 16$ samples, i.e. the extreme case of the system design when $L = N/4$ (refer to Subsection 2.2.4 for details). The increased CIR length has been adopted to study the impact of various channel model orders on the estimators' performance.

As from the standpoint of channel estimation MSE, it does not matter whether the system is SISO or MIMO (due to adoption of the independent pilot pattern), we model the transmitter with only one Tx antenna. The number of Rx antennas is varied, depending on the experiment, from 1 to 8. More than one Rx antenna is modelled only to find out whether the recursive channel estimators can benefit, aggregating statistical information from different SLs.

The number of pilot subcarriers in the majority of experiments is set to the required minimum of $P = 16$, with the equispaced placement in the band, and the pilot periodicity factor is $\underline{P} = 1$, i.e. each transmitted block carries pilot symbols. Design of the Wiener filterbank for the robust 2D channel estimators is made for $\overline{\omega}_{\text{D}} = 0.025\pi$ (refer to Subsection 3.4.2 for details). 2D estimators experience delay of $M - 1$ blocks before the start of operation due to the necessity to fill up the filterbank memory.

The recursive estimator structures initiate at $\hat{\mathbf{T}}_{\underline{\tilde{\mathbf{h}}}\underline{\tilde{\mathbf{h}}}-1} = \iota\mathbf{I}$ (RMMSE) and $\hat{\mathbf{T}}_{\underline{\hat{\mathbf{h}}}\underline{\hat{\mathbf{h}}}-1} = \iota\mathbf{I}$ (RMMSE-MMSE), where the parameter $\iota = 10^4$. Initial FF setting for the adaptation algorithm described in Tab.3.1 is $\alpha_{\text{FF}}(-1) = 0.1$. Processing frame length is $I = 10$ blocks. Precision variables, υ and ν, are selected to be equal to 1% and 0.1% respectively.

The convergence behaviour of the recursive estimators is monitored by block-wise computation of the *normalised sum of square errors* (NSSE):

$$NSSE_m = \frac{1}{N N_{\text{tx}} N_{\text{rx}}} \sum_{i=0}^{N_{\text{tx}}-1} \sum_{j=0}^{N_{\text{rx}}-1} \left(\hat{\mathbf{h}}_m(j,i) - \mathbf{h}_m(j,i)\right)^{\text{H}} \left(\hat{\mathbf{h}}_m(j,i) - \mathbf{h}_m(j,i)\right) \tag{3.129}$$

that is averaged over 100 realisations of the channel response process.

The modelled channel is block-wise time-variant with the Doppler spectrum given by the Jakes model [118] and the maximum angular shift $\omega_D = 0.025\pi$, which represents the tolerable maximum according to (2.41). Four different multipath models have been adopted for simulation:

Ch.1) $K = L$ sample-spaced independent identically distributed (IID) components with the flat power distribution, defined in (2.14) by the factor $\alpha = 0$ (uniform PDP) and the corresponding rms delay spread $\tau_{rms} = 4.61B^{-1}$ (2.18);

Ch.2) 2-path ($K = 2$) sample-spaced model, which is equivalent to Ch.1 in the rms delay spread sense and is described by (2.22)-(2.25);

Ch.3) 2-path ($K = 2$) non-sample-spaced model, which is equivalent to Ch.1 in the rms delay spread sense and is described in Subsection 2.2.2.3;

Ch.4) $K = 3$ equipowered non-sample-spaced IID components with the non-sample-spaced excess delays $\tau_0 = 0$, $\tau_1 = 1.7B^{-1}$ and $\tau_2 = 4.4B^{-1}$ (illustrated in Fig.2.8 up to the 7th sample) and the corresponding rms delay spread $\tau_{rms} = 1.81B^{-1}$ (where B is the system bandwidth).

The modelled MIMO channel properties are in full accordance with the assumptions made in Subsection 2.3.1.

3.6.2 Performance of constrained channel estimators

MSE of CLS, CMMSE, CLS-MMSE and CMMSE-MMSE estimators can be computed using the formulas (3.13), (3.34), (3.97) and (3.104). Although these formulas yield theoretical results, it has been found that MSE values obtained by simulations coincide with those calculated analytically. Furthermore, all MSE results presented in Subsection 3.6.3 have been collected through the simulation experiments, and the reader can compare the corresponding scenarios to be convinced by the correctness of the statement.

All the MSE curves in this section are plotted versus SNR observed immediately at the pilot subcarriers, i.e. $SNR_p = \sigma_p^2 / \sigma_w^2$. This is because performance of the channel estimators depends only on the power allocated to the pilots, with no regard to the data symbol transmission in the other parts of the band. System performance dependence on the general SNR at the receiver input will be investigated in Chapter 4.

3.6.2.1 Intrablock estimators

Performance of the simplest CLS estimator (Fig.3.10) does not depend on the channel statistics. From (3.13) it is clearly seen that MSE is inversely proportional to SNR_p (hence the linear shape of the curve). This, however, cannot be said about the number of pilot subcarriers (P), in general. Assuming the pilot pattern to be modelled as proposed in Section 3.1, we plot MSE observed at $SNR_p = 25\text{dB}$ versus the number of pilot subcarriers, ranging from 16 to 64 in Fig.3.11. In contrast to the equispaced modes, corresponding to 16, 32 and 64 pilots, for which the linear MSE decline proportional to P always holds true, some of the other pilot patterns do not improve MSE with

respect to the configurations with the lesser pilot number. According to Fig.3.11, the "good" non-equispaced modes in the depicted range are those corresponding to 21, 30-31 and 58-63 pilot subcarriers. We do not claim optimality of the adopted pilot arrangement scheme in case of the non-equispaced modes. In fact, alternative (and apparently better) positioning is suggested in the work [91]. This example is just to show the necessity to design the system, so that $\mathrm{tr}(\mathbf{S}) \to L/P$ and channel estimation MSE (3.13) could be kept as low as possible.

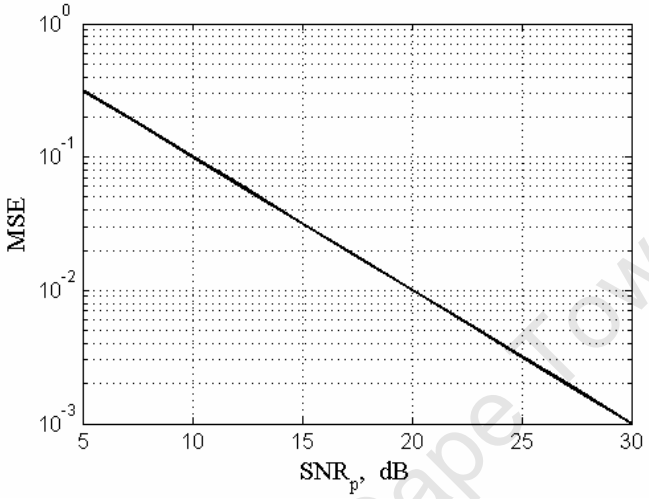


Fig.3.10. MSE of CLS estimator

Fig.3.12 and Fig.3.13 show MSE of several CMMSE estimator implementations, depending on the CIR correlation matrix adopted for the design, namely: $\overline{\mathbf{R}}_{\underline{\mathbf{hh}}} = \mathbf{R}_{\underline{\mathbf{hh}}}$ (optimal), $\overline{\mathbf{R}}_{\underline{\mathbf{hh}}} = \mathbf{R}_{\underline{\mathbf{hh}}} \circ \mathbf{I}$ (diagonal) and $\overline{\mathbf{R}}_{\underline{\mathbf{hh}}} = L^{-1}\mathbf{I}$ (robust). For all three variants, we assume the noise variance setting to be perfect, i.e. $\overline{SNR}_{\mathrm{p}} = SNR_{\mathrm{p}}$ in (3.33). In the sample-spaced channels (Ch.1 and Ch.2), optimal and diagonal estimators are identical as the CIR correlation matrix $\mathbf{R}_{\underline{\mathbf{hh}}}$ is diagonal due to statistical independence of the multipath components. This circumstance motivates system design with finer time granularity (wider bandwidth) and hence better multipath resolution, to benefit from the channel estimator implementation of lower complexity. In the non-sample-spaced channels (Ch.3 and Ch.4), the diagonal estimator is clearly outperformed by the optimal one. At the higher SNRs, it does not exhibit noticeable improvement with regard to the CLS scheme, whereas the optimal CMMSE maintains 9dB MSE gain in the entire SNR range for the 2-path channel model, irrespective whether the channel is sample-spaced or non-sample-spaced. The robust CMMSE design mode, which benefits from not relying on $\mathbf{R}_{\underline{\mathbf{hh}}}$, is, however, identical to CLS in terms of the MSE performance. Hence it is of no practical interest. Fig.3.12 shows that in the channels with a big number of multipath components (higher order models), the CMMSE estimator yields the same accuracy as CLS. To be more specific, increase of the channel model order from 2 (Ch.3) to 3 (Ch.4) leads to 1.7dB MSE loss (Fig.3.13). This loss tends to grow with the model order until no difference is observed between CMMSE and CLS.

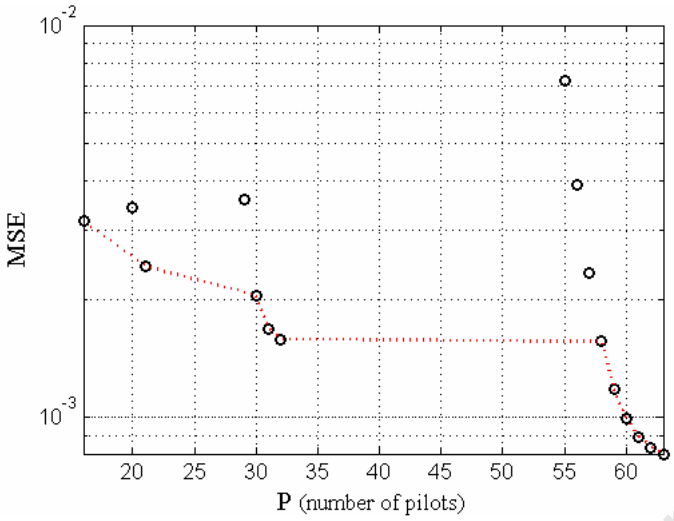


Fig.3.11. Dependence of CLS estimator's MSE (at $SNR_p = 25\text{dB}$) on the number of pilot subcarriers (dotted line encompasses reasonable modes)

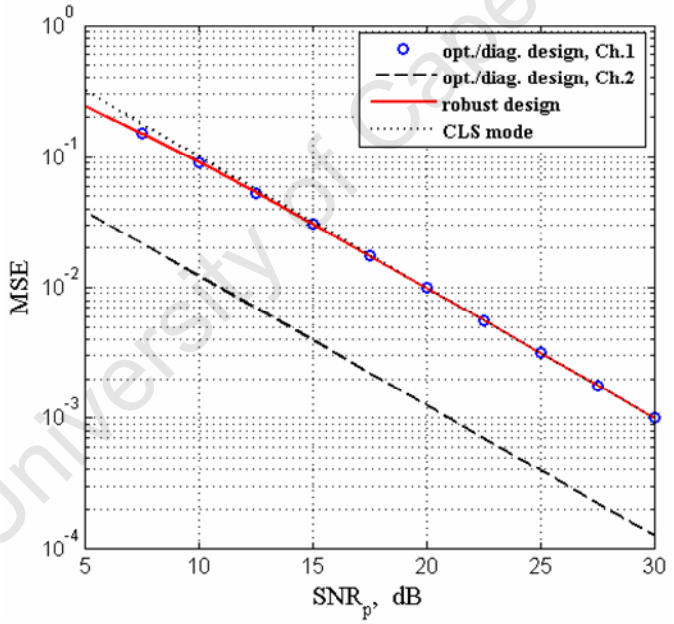


Fig.3.12. MSE of CMMSE estimator (Ch.1 and Ch.2)

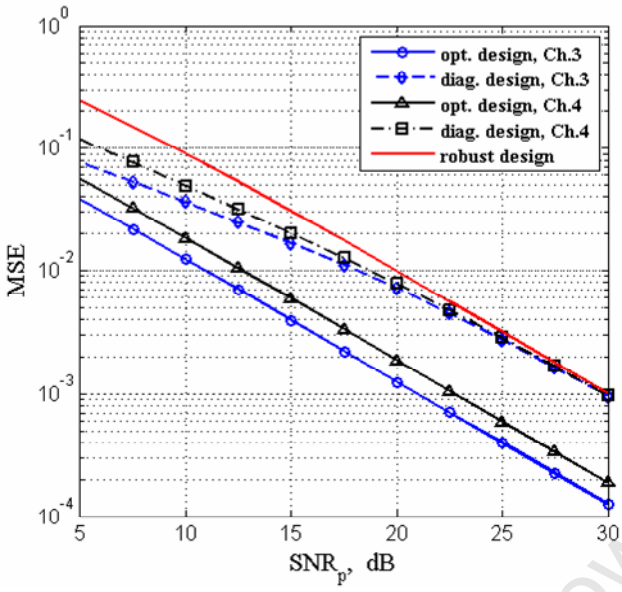


Fig.3.13. MSE of CMMSE estimator (Ch.3 and Ch.4)

Sensitivity of the CMMSE estimator with respect to the design setting of the CIR correlation matrix ($\overline{\mathbf{R}}_{\underline{\mathbf{hh}}}$) is shown in Fig.3.14, where the elements of $\overline{\mathbf{R}}_{\underline{\mathbf{hh}}}$ are affected by WGN to emulate inaccuracy of $\mathbf{R}_{\underline{\mathbf{hh}}}$ estimation (only the triangular part of $\overline{\mathbf{R}}_{\underline{\mathbf{hh}}}$ is considered in order not to violate the symmetry property of the correlation matrix), and $\overline{SNR}_{\mathrm{p}} = SNR_{\mathrm{p}} = 25\mathrm{dB}$. One can see that MMSE is achieved only if the estimates of $\mathbf{R}_{\underline{\mathbf{hh}}}$ elements are acquired having at least 63dB gain relative to the estimation noise variance. Thus, the requirement towards the design setting of the CIR correlation matrix is quite stringent, necessitating a more precise estimation of $\mathbf{R}_{\underline{\mathbf{hh}}}$.

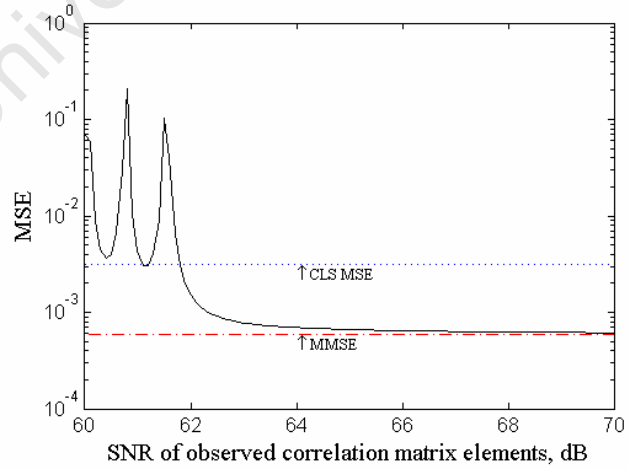
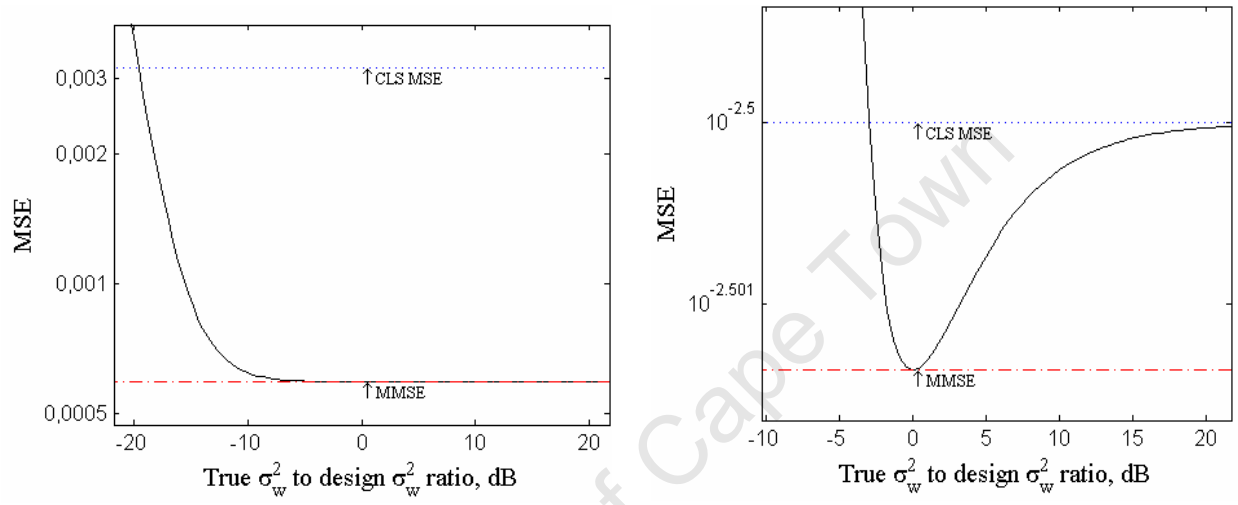


Fig.3.14. Impact of the noisy elements of the CIR correlation matrix, used for the CMMSE estimator design, on MSE (Ch.4, $SNR_{\mathrm{p}} = 25\mathrm{dB}$)

Analogously, we show in Fig.3.15 the importance of the accurate noise variance setting, $\overline{\sigma}_w^2$, used for the estimator design. The lower order channel model (Ch.4) allows for quite a wide range of $\overline{\sigma}_w^2$ values, with no remarkable deterioration of the estimator performance, whereas the higher order channel model (Ch.1) is more demanding towards $\overline{\sigma}_w^2$ selection (ideally $\overline{\sigma}_w^2 = \sigma_w^2$). Note that the numerical examples confirm correctness of the design constraint condition, specified in Subsection 3.3.4.3: the suboptimal CMMSE is no worse than the CLS estimator if it is designed for $\overline{\sigma}_w^2$ satisfying $0 \le \overline{\sigma}_w^2 \le 2\sigma_w^2$.



a) b)

Fig.3.15. Impact of the noise variance setting on MSE: a) Ch.4, b) Ch.1 ($SNR_p = 25$dB)

3.6.2.2 2D estimators

In the following examples, we study the robust 2D estimators, with the filterbank having lengths of $M = 10$, $M = 12$ and $M = 70$. The obvious disadvantages of the architectures with a longer filterbank memory are higher complexity and increased initialisation delay. These factors can negate the achievable performance gain in a practical implementation.

Fig.3.16 and Fig.3.17 show that the CLS-MMSE estimator with $M = 10$ loses about 2.2dB in comparison with the estimator with $M = 70$, which is not a very big difference. The CLS-MMSE estimator with $M = 70$ exhibits approximately 8.5dB, 16dB, 11.5dB and 10.5dB better MSE at lower SNRs than the intrablock CLS estimator, for Ch.1, Ch.2, Ch.3 and Ch.4, respectively. At higher SNRs, these gains tend to diminish to about 6dB, 15dB, 7.5dB and 7dB. The MSE gain reduction with the SNR growth is stronger manifested for the non-sample-spaced channels.

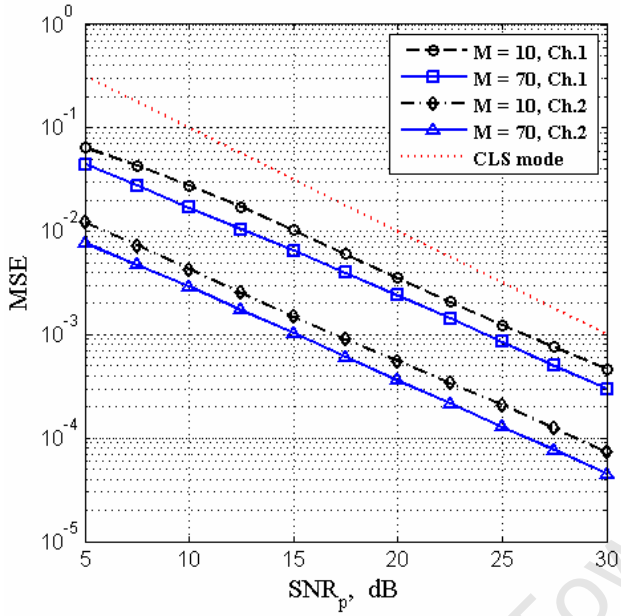


Fig.3.16. MSE of CLS-MMSE estimator (Ch.1 and Ch.2)

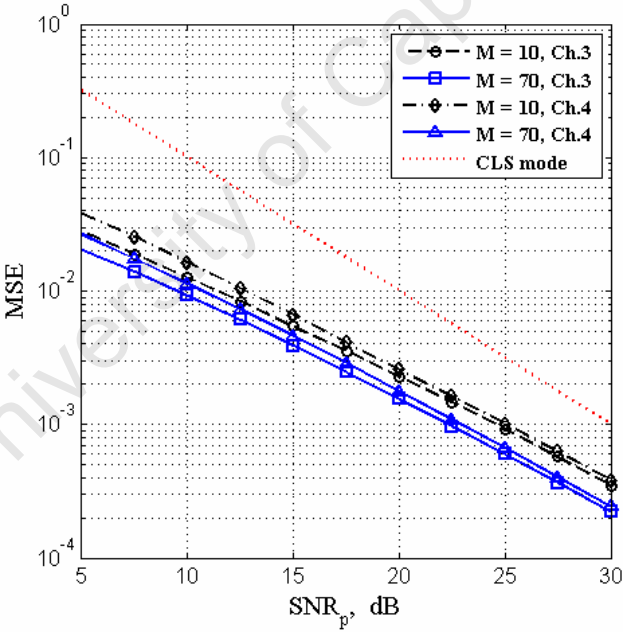


Fig.3.17. MSE of CLS-MMSE estimator (Ch.3 and Ch.4)

Impact of the length of the Wiener filters on MSE is illustrated by Fig.3.18, where the channel is modelled as Ch.4 and $SNR_p = 25\text{dB}$. MSE asymptotically tends to the minimum, corresponding to the infinite filter case ($M = \infty$). However, this trend is quite slow, and in practice it suffices to implement 20 to 50 FIR coefficients per filterbank branch, depending on the affordable complexity to achieve desired performance level.

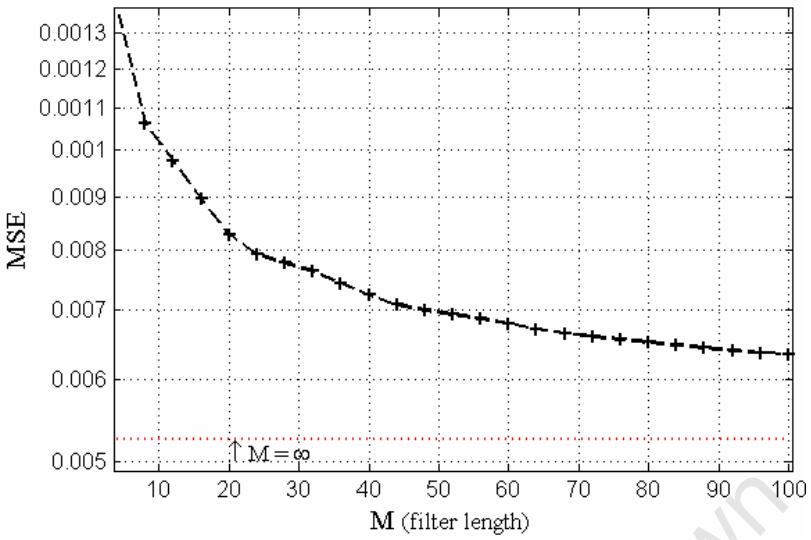


Fig.3.18. MSE dependence on filter length (Ch.4, $SNR_p = 25\text{dB}$)

It is also of interest to study the impact of Doppler variability of the channel response on the achievable estimator gain. For that purpose we model the maximum angular Doppler shift seen by the CLS-MMSE estimator, processing every $\underline{P}$ th block in the sequence, as $\tilde{\omega}_D = \underline{P}\omega_D = \overline{\omega}_D$ (refer to Subsection 3.4.2 for details) and let $\underline{P}$ vary from 1 to 40, i.e. $0.025\pi \le \tilde{\omega}_D \le \pi$. MSE is then calculated using the approximated equation (3.97) for the case of the finite-length filters and equation (3.83) for the infinite filter case. The results plotted in Fig.3.19 show that the increase of $\underline{P}$ and hence of the Doppler model order, defined by $\Omega^{-1} = \tilde{\omega}_D / \pi$, leads to the growth of MSE, which is quite rapid for the smaller values of $\underline{P}$. Nonetheless, it is interesting to point out that MSE of the CLS-MMSE estimator is smaller than that of the CLS estimator even for the maximum Doppler model order ($\Omega^{-1} = 1$). Finally, one should note that the difference between estimators of various filterbank lengths gradually vanishes with the growth of Ω^{-1} (larger pilot periodicity). These observations allow one to conclude that efficiency of the 2D channel estimation is strongly influenced by the rate, at which the pilot blocks are inserted into the sequence. System design, avoiding redundant symbol transmission and relying on CFR interpolation between successive blocks, limits applicability of the noise filtering methods based on the interblock channel correlation, and makes intrablock estimation approaches more preferable.

Fig.3.20 shows that in the sample-spaced channels, the CMMSE-MMSE estimator is unable to produce visible performance improvement over CLS-MMSE. In the non-sample-spaced channel scenarios (Fig.3.21), the observed MSE gain of the CMMSE-MMSE estimator with $M = 70$ is about 7.5dB for Ch.3 and 6dB for Ch.4 at higher SNRs. Similar to the CLS-MMSE case, the difference between the estimators with $M = 10$ and $M = 70$ is about 2.5dB. Absence of performance improvement for the sample-spaced channels allows one to conclude that the systems with high multipath resolution (e.g., multiband UWB) benefit fully from the CLS-MMSE estimator, which represents a much simpler implementation than CMMSE-MMSE.

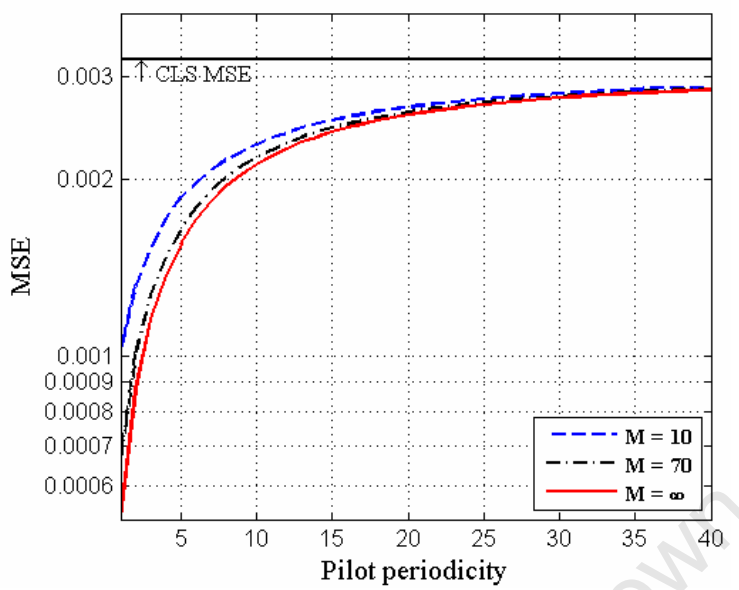


Fig.3.19. MSE dependence on the pilot periodicity coefficient $\underline{P}$

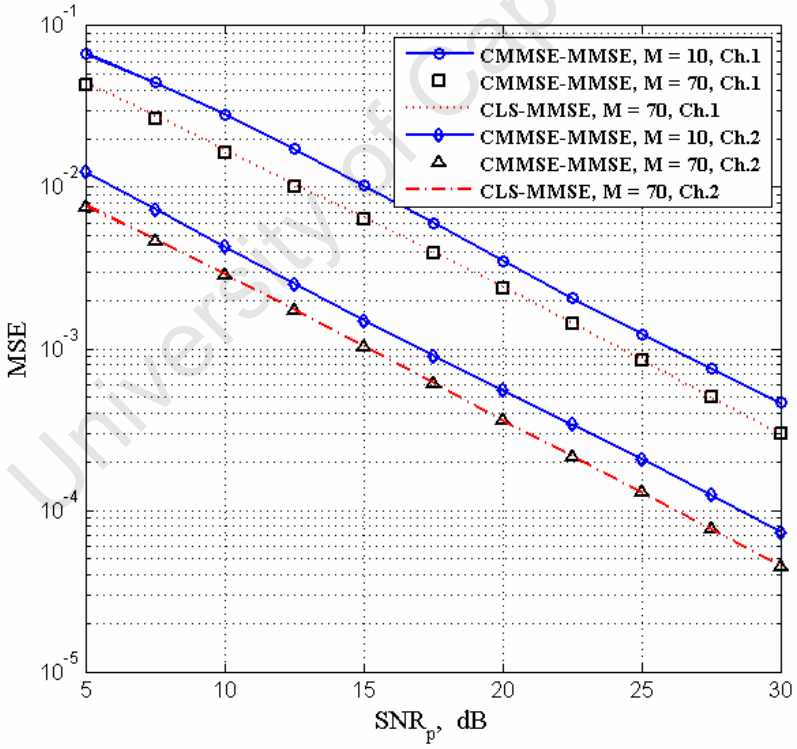


Fig.3.20. MSE of CMMSE-MMSE estimator (Ch.1 and Ch.2)

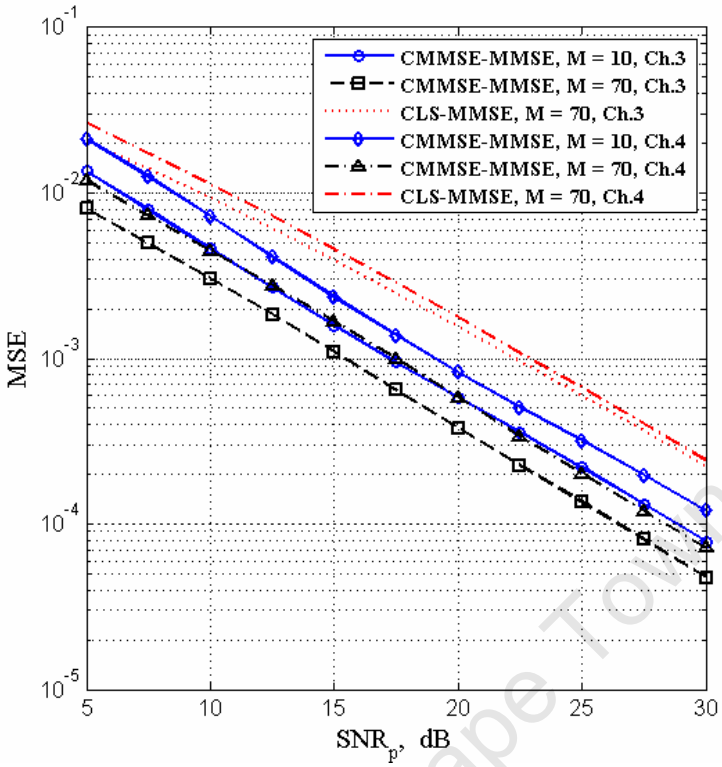


Fig.3.21. MSE of CMMSE-MMSE estimator (Ch.3 and Ch.4)

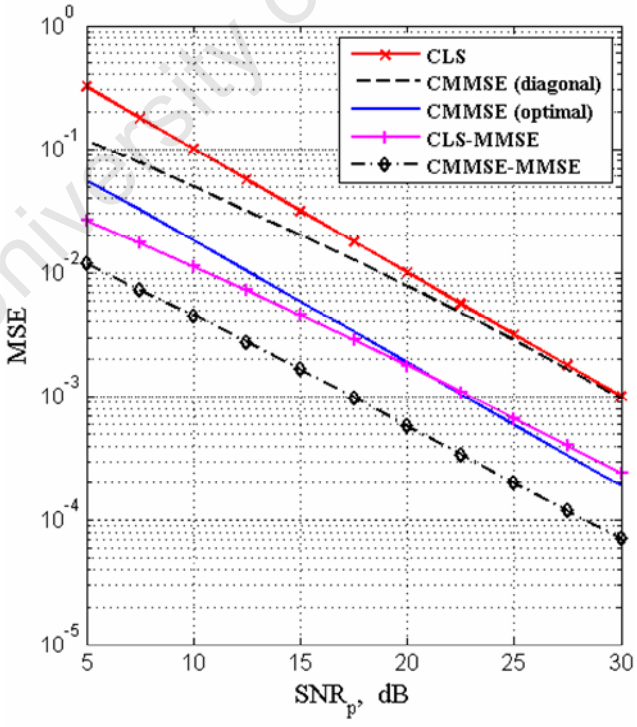


Fig.3.22. Comparison of estimators (Ch.4, $M = 70$)

The intrablock and 2D estimators (with $M = 70$) considered so far are compared in Fig.3.22, where the MSE, corresponding to the non-sample-spaced channel model Ch.4, is plotted. As it could be expected, the CMMSE-MMSE estimator achieves the best performance, whereas CLS is worse than others in the MSE sense. The difference between CMMSE-MMSE and CLS is about 13dB. It is obvious that the diagonal CMMSE implementation is not of much practical interest for the non-sample-spaced channels as it loses to the optimal CMMSE scheme up to 7dB at higher SNRs, which is as much as CLS. The CLS-MMSE estimator exhibits considerable MSE improvement with regard to CLS at lower SNRs and outperforms CMMSE, which is the best intrablock estimator, by approximately 3dB. Starting from $SNR_p = 21\text{dB}$ and higher, CMMSE demonstrates better performance than CLS-MMSE. However, it should be taken into account that the channel model under consideration (Ch.4) has only 3 independent multipath components. For the higher order channel models, the 2D estimators, relying on pilot symbol transmission in each block ($\underline{P} = 1$), would be absolutely better than CMMSE (e.g., refer to Fig.3.12 and Fig.3.16).

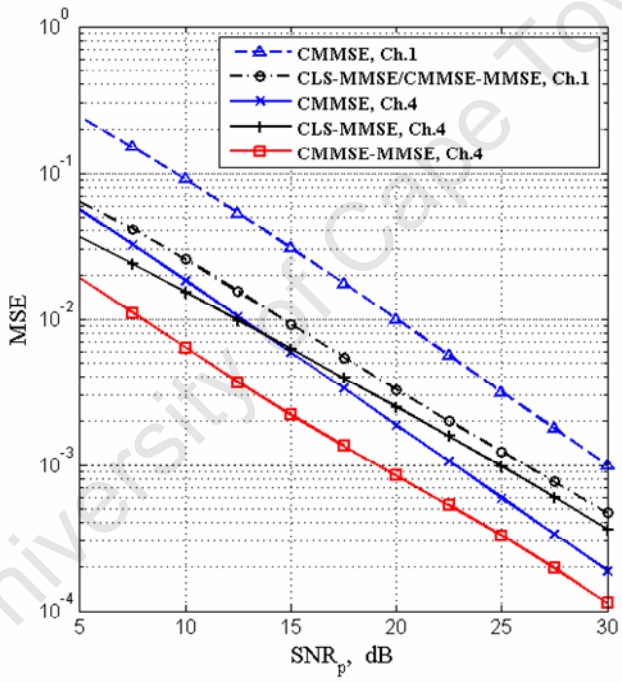


Fig.3.23. Comparison of estimators: CMMSE and CLS-MMSE are of the corresponding complexity ($M = 12$)

We continue comparing estimation algorithms in Fig.3.23, where we let CMMSE and CLS-MMSE estimators be analogous in the sense of the computational complexity inherent to their recursive implementations (RMMSE and RLS-MMSE, respectively) that implies the filterbank length being equal to $M = 12$ (refer to the numerical example in Section 3.5 for details). The CMMSE-MMSE estimator, considered in this scenario, has $M = 12$ too. In case of the sample-spaced channel with 16 independent multipath components (Ch.1), CLS-MMSE and CMMSE-MMSE demonstrate identical performance that is up to 4dB better than that of CMMSE in the observed SNR range. In the non-sample-spaced channel with 3 multipath components, the CMMSE estimator is able to improve MSE in

comparison with CLS-MMSE at higher SNRs, starting from $SNR_{\mathrm{p}} = 14\,\mathrm{dB}$ (it is less than the corresponding value in Fig.3.22 because of the MSE loss incurred by the short-memory filterbank design). Nonetheless, the CMMSE-MMSE algorithm always insures better MSE than CMMSE with approximately 2.5dB gain at higher SNRs.

3.6.3 Performance of recursive channel estimators

Most optimal and suboptimal channel estimators, considered in the previous subsection, rely on the assumption that all the necessary channel correlation parameters are known and preset. However, in practice, acquisition of these parameters often necessitates an implementation of the appropriate recursive mechanism at the receiver. It is obvious that performance of the estimator, relying on the estimated channel statistics, will always be poorer than that of the estimator based on the known correlation parameters. MSE of the converging recursive algorithm should tend in the steady state to MSE of its optimal counterpart having the fixed design. Hence we first analyse convergence behaviour of the developed recursive estimators at the initialisation stage and then evaluate their steady-state MSE performance by simulating transmission of 15000 OFDM blocks through the single-antenna system and Ch.4 channel model.

3.6.3.1 Initialisation of recursive estimators

Fig.3.24 and Fig.3.25 show the average normalised sum of square errors (ANSSE) of the RMMSE estimator, processing the first 150 transmitted blocks. The precision matrix's FF α_{FF} is fixed for the duration of simulation and equal to 0.01 (Fig.3.24) or 0.001 (Fig.3.25). The smaller α_{FF} magnitude implies slower convergence. In the steady state ($m \to \infty$), ANSSE approaches MSE of the optimal CMMSE estimator in both scenarios. The example with different fixed α_{FF} values shows that it is desirable to have a variable FF. Setting a relatively large FF would allow quickening estimator convergence at the initialisation stage if the channel dynamics are sufficient for computing a rough estimate of the CIR-plus-noise precision matrix (3.54). Later on the FF value can be lowered to smooth out the estimate of the precision matrix. In the MSE performance evaluation we show the advantages of the FF adaptation algorithm, proposed in Subsection 3.3.2.3 to meet this objective.

Comparing ANSSE learning curves of the estimators processing single and multiple SLs, it should be noted that averaging of the precision matrix estimate across several SLs improves convergence, particularly for small FF settings (Fig.3.25). As more SLs are aggregated, the faster the performance of the RMMSE algorithm approaches that of CMMSE.

One of the key advantages of the considered RMMSE architecture is the possibility to disable the smoothing module, forcing the estimator to operate in the robust CLS mode (refer to Subsection 3.3.2 for details), if accurate CIR-plus-noise statistics are not yet available at the receiver. In order to determine the optimal tolerance factor in the mode switching equation (3.29), we simulated initialisation of the system with the Ch.3 channel model. In this particular case of study, the modelled maximum angular Doppler shift is selected as $\omega_{\mathrm{D}} = 0.25\pi$, to allow for a

more rapid estimator convergence. FF value for computation of the cross-correlation in (3.29) is set to 0.1, to allow for a faster reaction of the system to changes in the channel statistics.

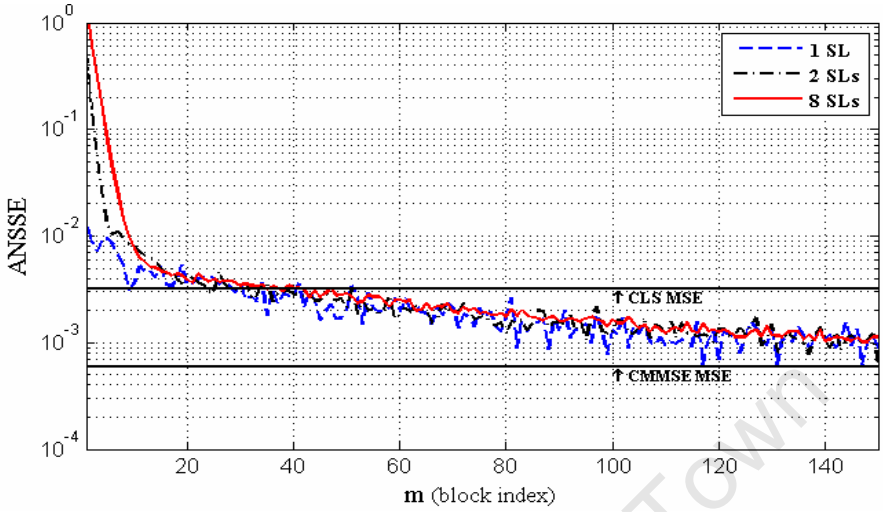


Fig.3.24. Learning curve of RMMSE estimator ($\alpha_{FF} = 0.01$, $SNR_p = 25\text{dB}$)

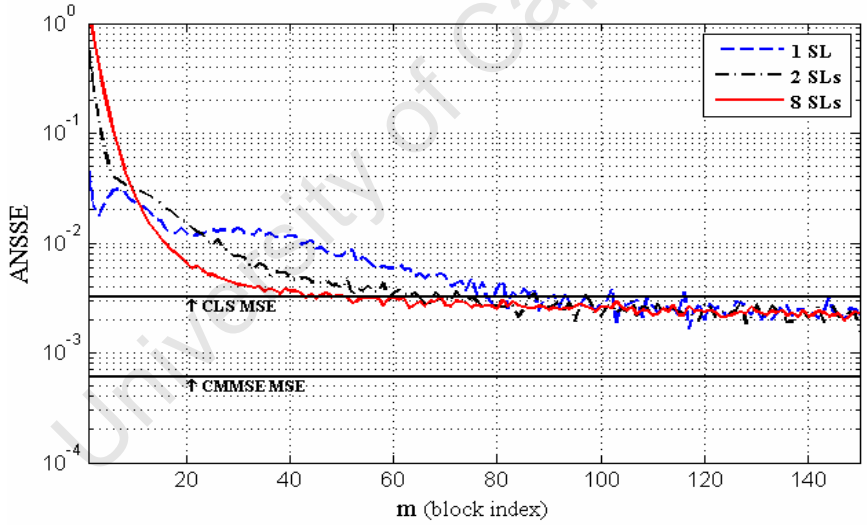


Fig.3.25. Learning curve of RMMSE estimator ($\alpha_{FF} = 0.001$, $SNR_p = 25\text{dB}$)

One can see from Fig.3.26 that the higher tolerance factor (ζ) values allow for early switching to the recursive mode, leading to occurrence of big error magnitudes inherent to the RMMSE initialisation. Lower ζ settings yield later triggering of switching to the recursive mode, as well as frequent returns to the CLS mode at NSSE surges. Both these factors limit the estimator's accuracy as the MSE is biased to the upper bound (CLS MSE in Fig.3.26) in the steady state. It has been found that the best value for ζ is about 1.4, i.e. 40% tolerance in excess of the nominal value of trace of the correlation matrix (3.28).

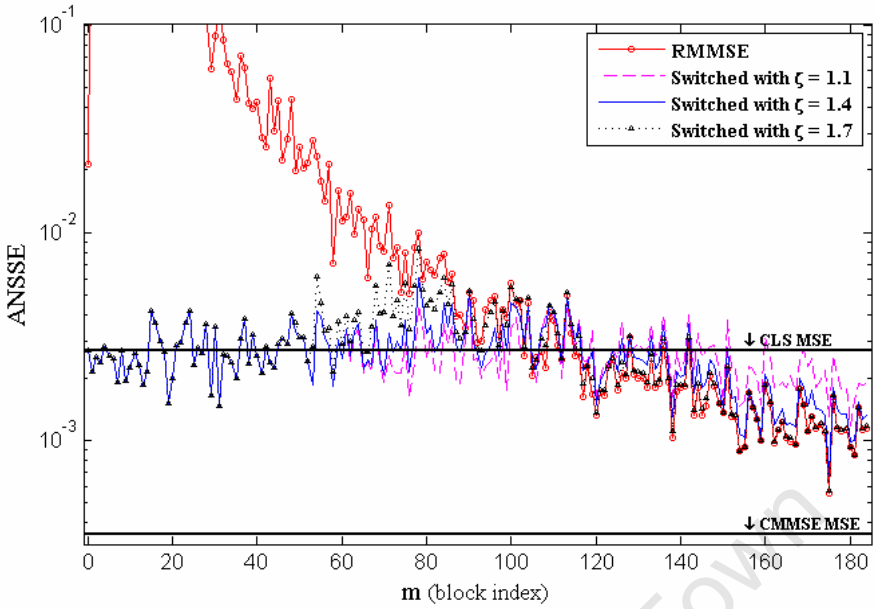


Fig.3.26. Impact of the tolerance factor (ζ) on operational mode (CLS/RMMSE) switching at RMMSE initialisation (Ch.3, $\alpha_{FF} = 0.01$, $SNR_p = 25\text{dB}$)

Presence of the FIR filterbank in the 2D estimators implies a mandatory delay of $M-1$ blocks, during which the estimator has to operate in the CLS mode (first 11 blocks in Fig.3.27 and Fig.3.28, as we consider an example of $M = 12$). The convergence behaviour of the RLS-MMSE estimator (Fig.3.27) is quite different from that of RMMSE. At the initialisation stage, there are frequent ANSSE surges of high magnitude. However, after processing about 100 blocks, the learning curve sticks with the optimal MSE level corresponding to the CLS-MMSE estimator.

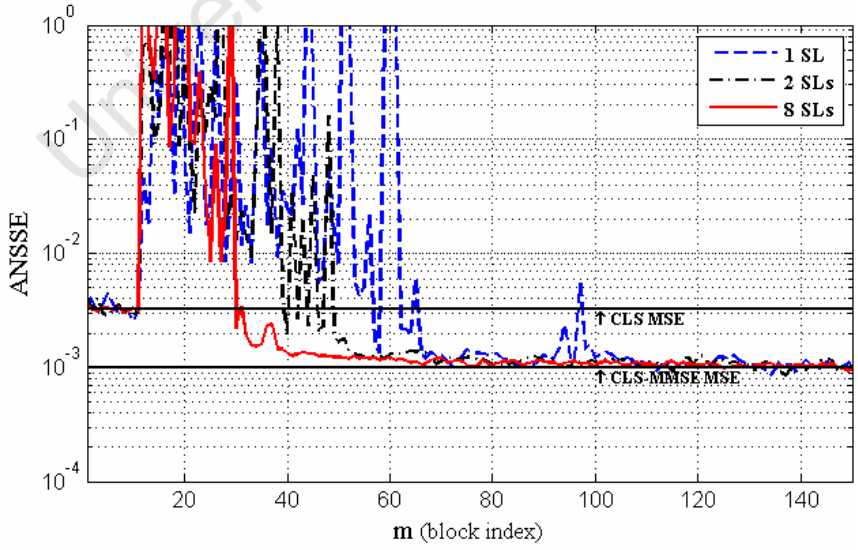


Fig.3.27. Learning curve of RLS-MMSE estimator ($M = 12$, $\alpha_{FF} = 0.01$)

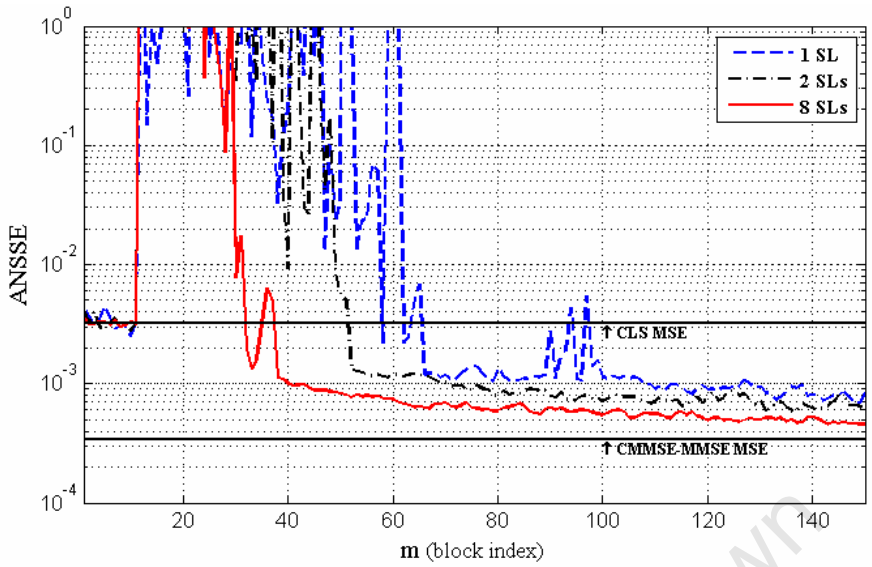


Fig.3.28. Learning curve of RMMSE-MMSE estimator ($M = 12$, $\alpha_{FF} = 0.01$, $SNR_p = 25\text{dB}$)

Convergence of the RMMSE-MMSE estimator is to a considerable degree influenced by the RLS-MMSE estimator, which constitutes the first two modules in the RMMSE-MMSE architecture (Fig.3.8); hence the similar look of the learning curve in Fig.3.28. However, ANSSE tends to the lower theoretical bound on MSE much slower than in the RLS-MMSE case. This could be explained by the slow convergence of the intrablock smoothing module, which has the same implementation as in the RMMSE scheme (refer to the RMMSE learning curve in Fig.3.24 for comparison).

Impact of statistics aggregation from multiple SLs on the convergence behaviour is stronger manifested for the 2D estimators. The RLS-MMSE learning curve (Fig.3.27) exhibits transient sections up to the 100th block in the single-antenna receiver system, whereas increasing SL number to 8 shortens initialisation interval to the first 40 blocks. The same observations hold for the RMMSE-MMSE estimator (Fig.3.28).

3.6.3.2 Steady-state performance of recursive estimators

Computation of the channel estimation MSE in the simulated examples excludes the estimator initialisation period, which is characterised by high-magnitude errors, in order to minimise the MSE bias due to the limited simulation duration. In our analysis, we consider scenarios with various preset FF values, as well as the proposed FF adaptation algorithm.

Fig.3.29 shows MSE of the RMMSE estimator, which is designed based on the known SNR_p at the receiver. It can be seen that performance of the RMMSE estimator with the precision matrix's FF adaptation is very close to the optimum MMSE performance (only 0.3dB loss with regard to CMMSE). On the other hand, a preset FF value might not be enough to ensure sufficient averaging memory for the CIR correlation matrix estimation that in turn leads to the MSE growth (about 13.5dB for the case of $\alpha_{FF} = 0.1$, and 1.5dB for the case of $\alpha_{FF} = 0.01$). It should be noted that even a small update interval ($I = 10$ blocks) in the adaptation algorithm allows for quite accurate FF

computation. Fig.3.30 shows FF variation along the simulated block sequence. On the post-initialisation stage the FF update interval can be set longer, which would potentially allow production of smaller FF values, leading to the averaging memory increase, which is necessary for correlation estimation in the channels with minor Doppler variation (e.g., fixed wireless access channels).

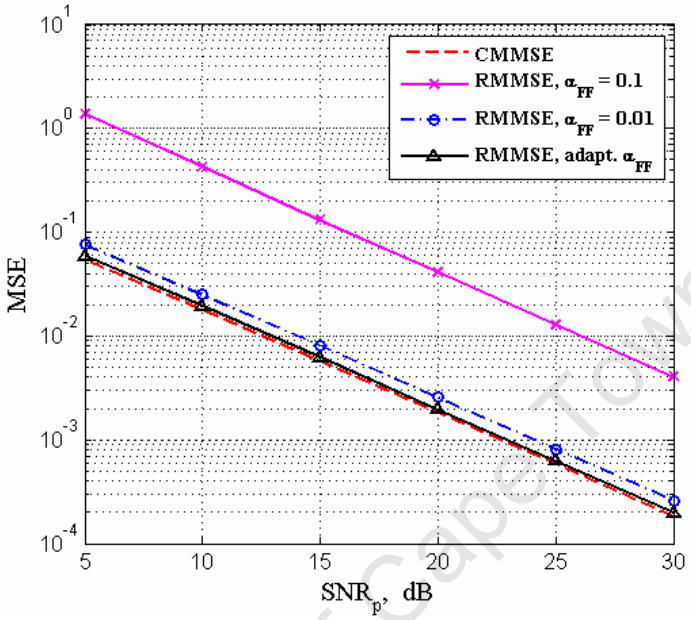


Fig.3.29. MSE of RMMSE estimator (known SNR_p at the receiver)

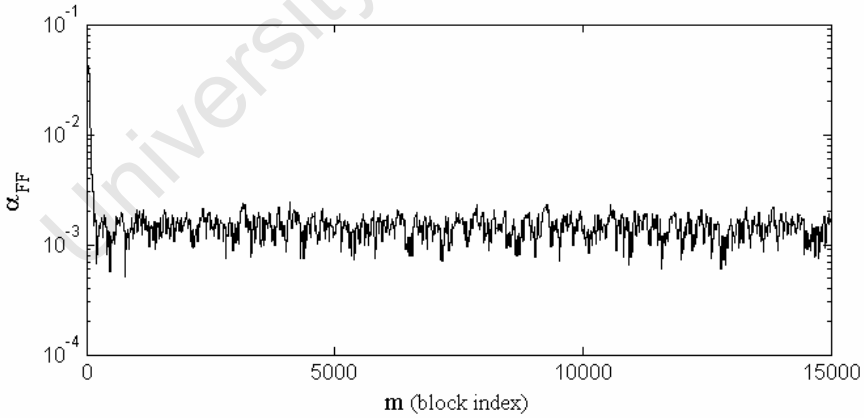


Fig.3.30. Forgetting factor adaptation process

Fig.3.31 illustrates the importance of information about SNR at the receiver input. The estimate of the noise variance for the *m*th block in the simulated transmit sequence, $\hat{\sigma}^2_{wm}$ ($\hat{SNR}_{pm} = \sigma^2_p / \hat{\sigma}^2_{wm}$), has been obtained using the exponential averaging with the short averaging memory (FF set to 0.1). Noise power measurement at single VC or at a pair of VCs has been found satisfactory for the RMMSE estimator (only about 1dB and 0.5dB accuracy loss,

respectively). Larger number of VCs, as well as increased averaging memory (smaller FF) improves performance by shifting the MSE curve closer to the one characterised by the known SNR_p.

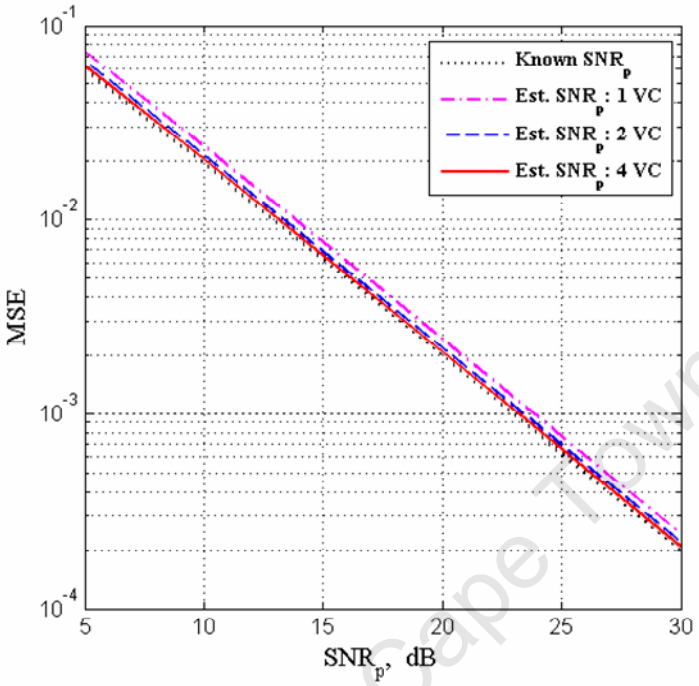


Fig.3.31. MSE of RMMSE estimator (estimated SNR_p at the receiver)

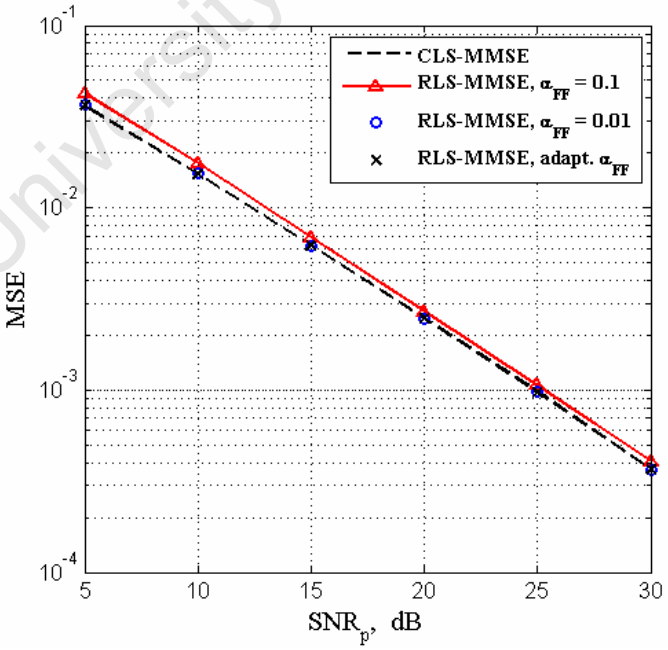


Fig.3.32. MSE of RLS-MMSE estimator (known SNR_p at the receiver)

MSE of the RLS-MMSE estimator (Fig.3.32) has been found less sensitive to the FF magnitude than that of the RMMSE algorithm. Setting $\alpha_{FF} = 0.1$ ensures less than 1dB accuracy loss with regard to the CLS-MMSE estimator. Smaller preset FF value ($\alpha_{FF} = 0.01$), as well as FF adaptation, yields the same performance in the steady state as CLS-MMSE.

In contrast to RLS-MMSE, the RMMSE-MMSE algorithm exhibits considerable MSE loss in the lower SNR range (Fig.3.33). This loss is obviously caused by operation of the intrablock smoothing module, which is the last one in the cascade (after the RLS-MMSE part in Fig.3.8). Unlike the RMMSE scheme, the smoothing module relies not only on the estimated precision matrix $\hat{\mathbf{T}}_{\underline{\hat{h}}\underline{\hat{h}}m}$, but also on the estimated error correlation matrix $\hat{\mathbf{R}}_{\underline{\varepsilon}\underline{\varepsilon}m}$ at the filterbank output. Using FF adaptation in the RMMSE-MMSE estimator allows reduction of MSE to 1dB at higher SNRs.

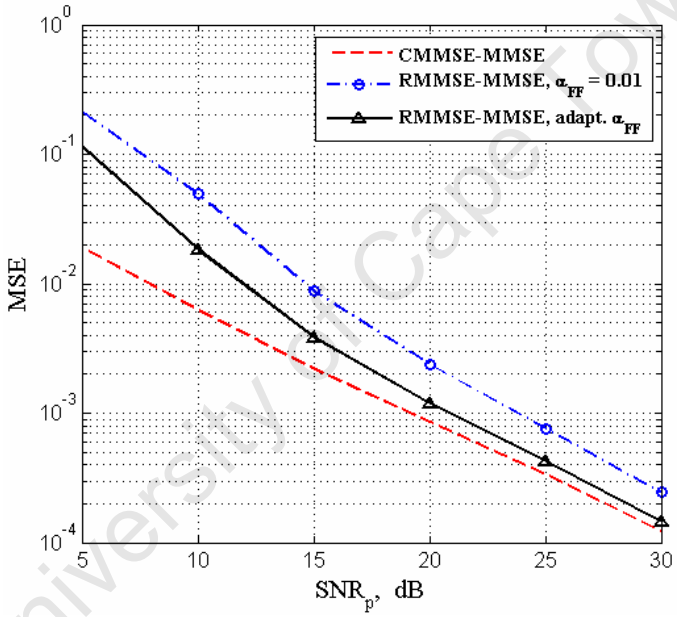


Fig.3.33. MSE of RMMSE-MMSE estimator (known SNR_p at the receiver)

In analogy to the estimators with the known channel statistics (Fig.3.23), we compare the steady-state MSE performance of the considered recursive algorithms here. One can see from Fig.3.34 that for $SNR_p \le 10\text{dB}$ the best estimator is RLS-MMSE. At $SNR_p > 15\text{dB}$, it is, however, outperformed by the RMMSE estimator with the identical computational complexity. The RMMSE-MMSE algorithm is seen to combine the advantages of RMMSE and RLS-MMSE in the medium-to-high SNR range. It demonstrates 10dB gain in comparison with the simplest CLS estimator at $SNR_p = 30\text{dB}$. Simulation results shown in Fig.3.34 lead to a conclusion that the last (smoothing) module in the RMMSE-MMSE architecture should be enabled only when $SNR_p > 15\text{dB}$ and the system is found in the steady state, i.e. reliable estimates of the channel correlation statistics have been produced. In all other cases the

RLS-MMSE mode is more preferable. Furthermore, during the initialisation period (100 first blocks in Fig.3.27) the filterbank module should also be disabled letting the estimator work in the CLS mode.

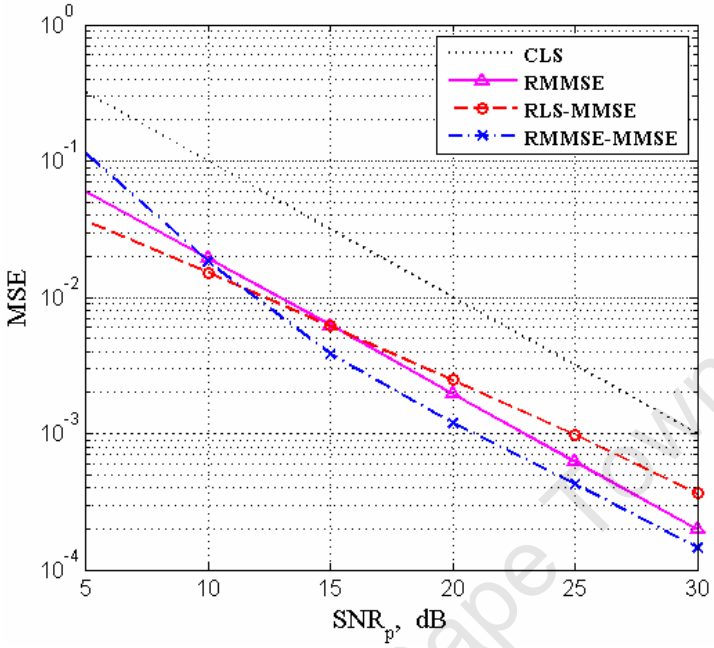


Fig.3.34. Comparison of recursive estimators: RMMSE and RLS-MMSE have identical complexity ($M = 12$)

3.7 Summary

In this chapter, we have developed low-complexity linear channel estimators for the pilot-assisted MIMO-OFDM system that have been categorised into the two major classes: one-dimensional intrablock and two-dimensional. The estimators have been presented in the optimal and suboptimal (robust) variants, which can be adopted, depending on the channel properties and availability of the channel correlation information at the receiver. The key underlying principles towards the reduced complexity design are transform-domain processing (use of the FFT/IFFT engines) and separable filtering (modular structures). Another important ingredient of the practical estimator implementation is the recursive mechanisms of the channel correlation estimation.

The ultimate reduced-complexity recursive estimator structures (Fig.3.2, Fig.3.7 and Fig.3.8) benefit from the cascaded design. Depending on the operation conditions (initialisation or steady state, high or low SNR) some of the filtering modules in the cascade can be enabled or disabled to achieve the best performance, e.g., the smoothing module of the RMMSE estimator (Fig.3.2) and RMMSE-MMSE estimator (Fig.3.8) can be disconnected if the accurate estimates of the CIR-plus-noise/error precision matrix have not yet been obtained. The filterbank module of the RLS-MMSE scheme (Fig.3.7) has initialisation latency, which is necessary to fill up the filter memory and to

produce accurate estimates of the channel PDP. During the filterbank initialisation the estimator can operate in the CLS mode by just unplugging the filterbank.

Based on the numerical analysis in Section 3.6, we highlight the advantages and drawbacks of different practical realisable estimators, described in this chapter, and suggest their applications. The corresponding summary is presented in Tab.3.5.

Tab.3.5. Summary of pilot-assisted channel estimators

Estimator	Advantages	Disadvantages	Application
CLS	1) Lowest complexity 2) Does not rely on channel statistics and SNR information	Poor MSE performance	1) Mobile channels with non-stationary interblock CIR variation 2) High-dimensional MIMO systems (complexity limitations incurred by a big number of SLs)
RMMSE (optimal design)	1) High MSE gains in the channels with a few multipath components 2) Performance is independent of pilot periodicity	1) Relies on CIR-plus-noise precision matrix and SNR information 2) Performs identically to CLS in the channels with many multipath components	1) Mobile channels with a few multipath components, constant intrablock CIR correlation and potential non-stationarities in interblock CIR variation 2) Systems with large pilot periodicity (applicable in the fixed radio channels)
RMMSE (diagonal design)	1) The same as for optimally designed RMMSE 2) Reduced complexity due to diagonal implementation of the CIR-plus-noise precision matrix	1) Relies on channel PDP and SNR information 2) MMSE is not achieved in the non-sample-spaced channels because of the non-optimal design that leads to CLS performance in the worst case 3) Applicable only in case of the equispaced pilot pattern	1) The same as for optimally designed RMMSE 2) Limited to the sample-spaced multipath channels (e.g., in multiband UWB systems)
RLS-MMSE	1) High MSE gain at lower SNRs for arbitrary channels 2) Fast convergence at initialisation in comparison with other estimators	1) Relies on channel PDP and SNR information 2) MSE loss due to the finite filterbank memory 3) Complexity and initialisation delay increase with filter length	1) Channels with constant PDP and stationary interblock CIR variation 2) Systems with small pilot periodicity 3) Systems with multiple SLs (estimator convergence improvement by statistics aggregation)
RMMSE-MMSE	The best performance at medium-to-high SNRs in comparison with other estimators	1) The same as for RLS-MMSE 2) Relies on CIR-plus-error precision matrix, CIR correlation matrix and SNR information 3) Additional complexity due to intrablock smoothing module (higher than that of RMMSE) 4) Poor performance at low SNRs 5) Slow convergence at initialisation 6) Performs identically to RLS-MMSE in the channels with many multipath components	1) Fixed wireless access channels with a few multipath components 2) Systems with small pilot periodicity 3) Low-dimensional MIMO systems (complexity implications)

4 Optimal Pilot Design for Linear SM-MIMO-OFDM Systems

Spectral efficiency of the wireless communication system is challenged by the receiver's ability to accurately identify data symbols at the channel output. In Chapters 2 and 3, we have established that in practical scenarios the detector's performance is always limited, due to the presence of white Gaussian noise (WGN) and channel estimation errors. Accuracy of the channel state information (CSI) retrieved at the receiver is of primary importance as the detector operation relies on CSI to extract data symbols from the WGN background. It is obvious that excessive channel estimation errors will bring to naught the detector's efforts, so that even an optimal detector design will appear useless. It has been found in Chapter 3 that more accurate estimates of the channel frequency response (CFR) in the pilot-assisted OFDM system can be obtained by either increasing the number of subcarriers dedicated to pilots in the band or assigning more power to the pilot symbols, to achieve higher SNR at the pilot subcarriers. Both of these measures should be treated with caution as in the former case larger amount of pilots lowers the channel capacity available for data transmission, whereas in the latter case the total transmit power constraint implies sacrificing the power budget of the data subcarriers to that of the pilot ones. A need to improve channel estimation and detection performance concurrently leads to the dilemma of the optimal power assignment and spectrum distribution between data and pilot subcarriers. One should note that placement of the pilot subcarriers in the band is also regarded as part of the system optimisation. However, unlike the number and power of pilot symbols, this problem is of the local type and can be solved, taking into account only performance of the channel estimator. In Subsection 3.3.4, it has been shown that the equispaced arrangement of the equipowered pilot subcarriers minimises MSE of the generally designed intrablock estimator. The same rule applies to the linear 2D estimators with the modular structure (Section 3.4) as they represent extensions to the intrablock one and hence are dependent on the primary filtering modules in the MSE sense.

In this chapter, we consider in detail the transmit parameters, which influence overall system performance as seen at the ouput of the detector (detection MSE) and demodulator (symbol error rate). Section 4.1 discusses what should be the optimal number of pilot subcarriers and Tx antennas in the system with continuously and periodically transmitted pilot blocks. The problem of optimal power allocation between the pilot and data symbols is formulated and analytically solved for various types of channel estimators in Section 4.2. The derived analytical results are subsequently validated through several numerical examples in Section 4.3.

4.1 Optimal value and constraint on the number of pilot subcarriers and transmit antennas

Consider a pilot-assisted SM-MIMO-OFDM system with N subcarriers and the *independent* pilot arrangement across Tx antennas (refer to Section 3.1 for details). Let the number of pilot subcarriers associated with a single Tx antenna be equal to P, and the number of Tx antennas be equal to N_{tx}. Assume that the pilot-bearing blocks are transmitted with the period of $\underline{P}$ blocks where $\underline{P}$ is also known as the *pilot periodicity* coefficient (see Subsection 3.4.2). Decoupling of channel estimation at different spatial layers (SLs) necessitates placement of virtual symbols (VSs) at the selected pilot subcarrier by other Tx antennas (as explained in Section 3.1), reducing the number of symbols available for data transmission per pilot block to $N_{\text{tx}}(N - P N_{\text{tx}})$, in contrast to the non-pilot blocks carrying up to $N_{\text{tx}} N$ data symbols.

It has been shown in Section 3.1 that the number of pilot subcarriers per Tx antenna has to be no less than the anticipated channel impulse response (CIR) length L for guaranteed error-free CFR recovery in the entire band, in the absence of noise and information about the channel statistics. The related work analysis [94] established that setting $P = L$ is not only optimal in terms of the bandwidth efficiency, but also in the sense of maximising the lower bound on the average ergodic capacity.

It is clear that the capacity of the theoretical SM-MIMO-OFDM model without the in-band pilot transmission is proportional to the number of Tx antennas. However, in the case of the pilot-assisted system that is not true as a certain portion of bandwidth is consumed in training the channel estimators. Moreover, the training-dedicated bandwidth increases with the number of Tx antennas because each Tx antenna is assigned its own set of the pilot subcarriers. This capacity-limiting trend is quantified by analysing the average number of symbols allocated for data transmission per block, being equal to:

$$D = \underline{P}^{-1} N_{\text{tx}}(N - P N_{\text{tx}}) + \underline{P}^{-1} N_{\text{tx}} N (\underline{P} - 1) = \underline{P}^{-1} N_{\text{tx}}(\underline{P} N - P N_{\text{tx}}). \tag{4.1}$$

D is a concave function of N_{tx} if $N_{\text{tx}} \le N/P$, and has the maximum at

$$N_{\text{tx,opt}} = \arg \max_{1 \le N_{\text{tx}} \le P^{-1} N} D = \begin{cases} N\underline{P}/(2P), & \text{if } \underline{P} \le 2 \\ N/P, & \text{if } \underline{P} > 2 \end{cases} \tag{4.2}$$

which is equal to

$$D_{\text{opt}} = \begin{cases} N^2 \underline{P}/(4P), & \text{if } \underline{P} \le 2 \\ N^2 (\underline{P} - 1)/(\underline{P} P), & \text{if } \underline{P} > 2 \end{cases}. \tag{4.3}$$

Note that $N_{\text{tx,opt}}$ in (4.2) can take non-integer values in general. Hence the actual optimal number of Tx antennas is defined as $\lfloor N_{\text{tx,opt}} \rfloor$, where $\lfloor . \rfloor$ denotes rounding to the smaller integer.

The constraint on the number of Tx antennas, $1 \le N_{tx} \le N/P$, originates from inequality (3.4), meaning that the total number of pilot symbols per block, $P N_{tx}$, cannot exceed the number of subcarriers N. This leads to the limitation of the optimal MIMO capacity, given by the second equality variant in (4.3). This motivates the use of another training structure for the MIMO-OFDM systems with $\underline{P} > 2$ that is different from periodical transmission of the single pilot-bearing blocks with the independent pilot pattern, which is considered in this thesis. For the case of $\underline{P} \ge N_{tx}$, a possible option is the periodical bursts of the successive pilot blocks where each block in the burst carries a set of pilot subcarriers assigned to the individual Tx antenna. We, however, do not emphasise the pilot structure in the time-frequency grid as the optimal periodicity of the pilot block transmission depends on the Doppler statistics of the channel, SNR at the pilot subcarriers, as well as the adopted interblock channel response interpolation technique. The latter is not within the scope of the thesis as the main focus is on the optimal filtering methods, to retrieve CSI from the noise background in the absence of interpolation errors.

In Subsection 2.2.4, it has been discussed that a reasonable system design is possible under the condition of $L \le N/4 + 1 \approx N/4$. Assuming the optimal number of pilot subcarriers per Tx antenna in (4.2), i.e. $P = L \le N/4$, and $\underline{P} \le 2$, one can deduce that $N_{tx,opt} \ge 2\underline{P}$ and $D_{opt} \ge N\underline{P}$. Thus, for the largely dispersive channels (when $L = N/4$), optimal design is restricted to only two Tx antennas if $\underline{P} = 1$, and only four Tx antennas if $\underline{P} = 2$.

Tab.4.1 compares optimal capacity of the SM-MIMO-OFDM system, calculated according to (4.1) for different numbers of pilot symbols. One can see that increasing the pilot block period from 2 to 4 yields only 1.5-fold capacity growth, in contrast to the increase of the pilot periodicity from 1 to 2 that is characterised by doubling the capacity. This is due to the constraint on the number of Tx antennas, $N_{tx} \le N/P$, for $\underline{P} > 2$. It can also be seen that the data capacity is always inversely proportional to the number of pilot subcarriers that clearly conforms to equation (4.3).

Tab.4.1. Average number of data symbols per subcarrier in the pilot-assisted SM-MIMO-OFDM system

Pilot periodicity ($\underline{P}$)	**No. of pilots per Tx antenna** (P)	**Optimal no. of Tx antennas** ($N_{tx,opt}$)	**Average no. of data symbols per subcarrier** (D/N)
1	$N/10$	5	2.5
	$N/8$	4	2
	$N/6$	3	1.5
	$N/4$	2	1
2	$N/10$	10	5
	$N/8$	8	4
	$N/6$	6	3
	$N/4$	4	2
4	$N/10$	10	7.5
	$N/8$	8	6
	$N/6$	6	4.5
	$N/4$	4	3

4.2 Optimisation of pilot-to-data power ratio (PDR)

The optimal design of the SM-MIMO-OFDM system assumes that the power of all the pilot symbols is constant and equal to σ_{p}^2 (refer also to the definition in Section 3.1). Similarly, the average power of data symbols is also constant for all subcarriers and SLs, being equal to σ_{d}^2 (refer to the definition in Subsection 2.3.3). The equality of the transmit symbol power allows the use of *pilot-to-data power ratio* (PDR)

$$\kappa = \sigma_{\mathrm{p}}^2 / \sigma_{\mathrm{d}}^2 , \tag{4.4}$$

to quantify power distribution between training and data. Thus, PDR (4.4) is expected to have some optimal magnitude to yield the best performance (minimum error rate) of the pilot-assisted SM-MIMO-OFDM system.

In this section, we aim to derive an analytical solution of the optimal or, if not possible, the suboptimal PDR given the system design parameters (number of subcarriers, number of pilots, number of transmit antennas, multipath and Doppler model orders, SNR, etc.). Based on the channel estimation MSE description in Chapter 3, in Subsection 4.2.1 we determine the statistics of the error at the output of the SM-MIMO detector relying on the imperfect CSI (from the channel estimator), namely the normalised MSE (NMSE), which is chosen as a cost function for PDR optimisation. Derivations of the (sub)optimal PDR for the cases of various intrablock and 2D channel estimators are presented in Subsections 4.2.2 to 4.2.4.

4.2.1 Analysis of detection MSE in the presence of channel estimation errors

For the system performance analysis, we adopt the basic zero-forcing (ZF) linear detector. It has been shown in Subsections 2.3.3 and 2.4.3 that despite its inherent simplicity, this detector type guarantees good performance for the non-square MIMO systems (with $N_{\mathrm{rx}} > N_{\mathrm{tx}}$), approaching the optimal linear detector, which is based on the minimum MSE criterion. The output of the ZF detector is described by equation (2.48),

$$\hat{\mathbf{x}}_{n,m} = \hat{\mathbf{H}}_{n,m}^{+} \mathbf{y}_{n,m} , \tag{4.5}$$

where $\hat{\mathbf{x}}_{n,m}$ denotes the vector of estimated data symbols at the input of N_{tx} Tx antennas on the nth subcarrier inside the mth block; $\mathbf{y}_{n,m}$ is the vector of symbols at the output of N_{rx} Rx antennas on the nth subcarrier inside the mth block; and the matrix $\hat{\mathbf{H}}_{n,m}$, containing the MIMO CFR estimates, is used instead of the true CSI matrix $\mathbf{H}_{n,m}$.

The entries of $\hat{\mathbf{x}}_{n,m}$ (4.5) are affected by WGN with variance σ_{w}^2, as well as the channel estimation errors concealed in $\hat{\mathbf{H}}_{n,m}$. The data symbol estimates, $\hat{\mathbf{x}}_{n,m}$, have to be subsequently demodulated (sliced on the constellation), restoring sent binary information. As a variant to improve performance, one can use the V-BLAST detector, relying on the decision feedback, instead of the parallel demodulation of the linearly estimated symbols (refer to Subsection 2.3.3.2 for details). Nonetheless, the decision feedback scheme is applied to process the "raw" estimates stacked in $\hat{\mathbf{x}}_{n,m}$ (4.5), hence at its initial stage (and therefore all subsequent iterations) the error probability

is directly influenced by the error variance at the output of the linear ZF detector. Thus, optimal power allocation between pilot and data symbols is not affected by the choice of the detection algorithm (linear ZF or V-BLAST with ZF weighting) and can be determined, taking into account minimisation of the error variance in (4.5).

Substitution of (2.47) into (4.5) yields the expression of the ZF detector's output that is approximated (see Appendix D.1) as

$$\hat{\mathbf{x}}_{n,m} = \mathbf{x}_{n,m} + \boldsymbol{\xi}_{n,m} \approx \mathbf{x}_{n,m} + \mathbf{H}^{+}_{n,m}(-\mathbf{E}_{n,m}\mathbf{x}_{n,m} + \mathbf{w}_{n,m}), \tag{4.6}$$

where $\mathbf{x}_{n,m}$ is the vector of the actual transmitted data symbols having the average power σ_{d}^2, $\mathbf{w}_{n,m}$ is the vector of WGN samples affecting received signal, $\boldsymbol{\xi}_{n,m}$ represents the zero-mean post-detection error term comprising of both the noise at the corresponding (*n*th) subcarrier and the CFR estimation error, and $\mathbf{E}_{n,m} = \hat{\mathbf{H}}_{n,m} - \mathbf{H}_{n,m} = [\varepsilon_{n,m}(j,i)]_{N_{\mathrm{rx}} \times N_{\mathrm{tx}}}$ is the matrix of the CFR estimation errors observed at the *n*th subcarrier.

In Appendix D.1 it is shown that the variance of $\boldsymbol{\xi}_{n,m}$, calculated for the given CFR realisation $\mathbf{H}_{n,m}$ and the channel estimation MSE $\sigma_{\varepsilon n}^2$, is equal to

$$\sigma_{\xi n}^2 = (\sigma_{\mathrm{d}}^2 \sigma_{\varepsilon n}^2 + N_{\mathrm{tx}}^{-1} \sigma_{\mathrm{w}}^2)\,\mathrm{tr}[(\mathbf{H}_{n,m}^{\mathrm{H}} \mathbf{H}_{n,m})^{-1}]. \tag{4.7}$$

From here one can see that the detection error variance is proportional to the channel estimation MSE. Hence an essential question arises on the impact of the selected channel estimator on the pilot-data power allocation. The optimal PDR, determined for one channel estimator type, will not be optimal for another estimator type.

We select the *normalised MSE* (NMSE) of detection at the *n*th subcarrier as a function for optimisation. (Alternatively it can be interpreted as the inverse of the *SNR gain* of the ZF detector output with respect to the receiver input.) However, instead of the exact CFR estimation error variance, $\sigma_{\varepsilon n}^2$, inherent to the *n*th subcarrier according to (4.7), we adopt its average, σ_{ε}^2, calculated over all N subcarriers. The corresponding average MSE expressions have been derived for a variety of channel estimators in Chapter 3. Furthermore, it has been shown that in the special case of the equispaced pilots and channel with the flat power-delay profile (PDP), CFR estimation variance of the optimal intrablock estimator (and hence of the 2D estimator too) is equal for all subcarriers in the spectrum (refer to Subsection 3.3.4 for details). For the general pilot arrangement we let detection NMSE be written as

$$\begin{aligned}\gamma &= SNR\, N_{\mathrm{tx}} \sigma_{\xi n}^2 \big/ \sigma_{\mathrm{d}}^2 \\ &= \mathrm{tr}[\mathrm{E}(\mathbf{H}_{n,m}^{\mathrm{H}} \mathbf{H}_{n,m})^{-1}]\, SNR\, N_{\mathrm{tx}}\, \sigma_{\mathrm{d}}^{-2} (\sigma_{\mathrm{d}}^2 \sigma_{\varepsilon}^2 + N_{\mathrm{tx}}^{-1} \sigma_{\mathrm{w}}^2) \\ &= c_{\mathrm{H}}\, SNR\, (1 + N_{\mathrm{tx}} SNR_{\mathrm{d}} \sigma_{\varepsilon}^2) \big/ SNR_{\mathrm{d}}\end{aligned}, \tag{4.8}$$

where *SNR* is the anticipated signal-to-noise power ratio at the receiver input, $SNR_{\mathrm{d}} = \sigma_{\mathrm{w}}^{-2} \sigma_{\mathrm{d}}^2$ is the SNR observed at the data subcarriers, $c_{\mathrm{H}} = \mathrm{tr}[\mathrm{E}(\mathbf{H}_{n,m}^{\mathrm{H}} \mathbf{H}_{n,m})^{-1}] > 0$, and the index n is omitted as it will not be involved in the subsequent derivations.

Given the noise variance, σ_{w}^2, which is constant across all subcarriers and Rx antennas, it is easy to show that $SNR_{\mathrm{d}} = \sigma_{\mathrm{w}}^{-2}\sigma_{\mathrm{d}}^2$ is related to the SNR observed at the receiver input according to the following equation:

$$SNR = \frac{P\sigma_{\mathrm{p}}^2 + (N - PN_{\mathrm{tx}})\sigma_{\mathrm{d}}^2}{N\sigma_{\mathrm{w}}^2} = \frac{SNR_{\mathrm{d}}[P\kappa + N - PN_{\mathrm{tx}}]}{N}, \tag{4.9}$$

where κ denotes the PDR (4.4). Hence we can express SNR_{d} and $SNR_{\mathrm{p}} = \sigma_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^2 = \kappa SNR_{\mathrm{d}}$ from (4.9) as

$$SNR_{\mathrm{d}} = \frac{SNR}{PN^{-1}\kappa + b}, \tag{4.10}$$

$$SNR_{\mathrm{p}} = \frac{SNR\,\kappa}{PN^{-1}\kappa + b}, \tag{4.11}$$

where

$$b = 1 - PN_{\mathrm{tx}}N^{-1} > 0. \tag{4.12}$$

Note here that SNR_{d} represents a decreasing function of PDR κ, whereas SNR_{p} is increasing. Recalling from Chapter 3 that the channel estimation MSE, σ_{ε}^2, decreases with the growth of SNR_{p}, one can treat σ_{ε}^2 as a decaying function of κ.

Inclusion of the normalising factor N_{tx} in the numerator of (4.8) is explained by the definition of *SNR* in (4.9), where the noise variance σ_{w}^2 is specified with respect to the superposition of signals from N_{tx} antennas at the channel output. Hence the corresponding average transmit symbol power, used for normalisation, is calculated as $SNR\,N_{\mathrm{tx}}\sigma_{\mathrm{w}}^2$.

In the following subsections, we optimise PDR for several cases of feasible low-complexity intrablock and 2D channel estimation algorithms described in Chapter 3. We emphasise the closed-form analytical expressions of the optimal magnitude and the upper bound (suboptimal magnitude) on PDR, due to their simplicity and hence direct applicability in contrast to the numerical optimisation methods. The resultant PDR equations can be applied to a system with arbitrary arrangement of the pilot subcarriers, operating in an arbitrary multipath channel.

It should be noted that although the considered theoretical channel estimators rely on the channel statistics and SNR known by the receiver, the same optimal PDR settings apply to their recursive counterparts as the recursive estimator converges to the theoretical one in the steady state.

4.2.2 PDR optimisation for the system with CLS channel estimator

The optimal PDR for the case of the CLS channel estimator is found by substituting MSE expression (3.13),

$$\sigma_{\tilde{\varepsilon}}^2 = s^{-1}SNR_{\mathrm{p}}^{-1} = s^{-1}\kappa^{-1}SNR_{\mathrm{d}}^{-1}, \tag{4.13}$$

where $s = 1/\mathrm{tr}[\mathbf{S}(0)]$, instead of σ_{ε}^2 in the detection NMSE formula (4.8):

$$\begin{aligned}\gamma &= c_{\mathrm{H}} SNR\,(1 + N_{\mathrm{tx}} SNR_{\mathrm{d}} \sigma_{\tilde{\varepsilon}}^2)\, SNR_{\mathrm{d}}^{-1} \\ &= c_{\mathrm{H}} (P N^{-1} \kappa + b)(1 + N_{\mathrm{tx}} s^{-1} \kappa^{-1}) \\ &= c_{\mathrm{H}} (P N^{-1} \kappa + N_{\mathrm{tx}} b\, s^{-1} \kappa^{-1} + P N_{\mathrm{tx}} N^{-1} s^{-1} + b)\,. \end{aligned} \tag{4.14}$$

It is easy to show that γ (4.14) is a convex function of κ ($\kappa > 0$) since

$$\frac{\partial^2 \gamma}{\partial \kappa^2} = \frac{2 c_{\mathrm{H}} N_{\mathrm{tx}} b}{s \kappa^3} > 0\,. \tag{4.15}$$

Hence the optimal PDR corresponds to the unique minimum of γ. Existence of this minimum can intuitively be explained as follows. SNR_{d} and σ_{ε}^2 in (4.8) have been defined as monotonically decreasing functions of PDR κ throughout (4.10) and (4.13). Therefore the detection NMSE γ, being proportional to the channel estimation MSE σ_{ε}^2, decreases with PDR enlargement, whereas a concurrent decrease of SNR_{d} leads to the growth of NMSE.

The optimal PDR $\kappa_{\mathrm{opt}} = \arg\min_{\kappa > 0} \gamma$ is determined by solving the quadratic equation

$$\frac{\partial \gamma}{\partial \kappa} = c_{\mathrm{H}} \left[\frac{P}{N} - \frac{N_{\mathrm{tx}} b}{s \kappa^2} \right] = 0\,, \tag{4.16}$$

yielding

$$\kappa_{\mathrm{opt}} = \sqrt{\frac{N_{\mathrm{tx}} N\, b}{s\, P}} = \sqrt{\frac{N_{\mathrm{tx}} (N - P N_{\mathrm{tx}})}{s\, P}}\,. \tag{4.17}$$

Note that the resultant PDR expression (4.17) does not depend on the channel properties, but only on the system configuration parameters.

4.2.3 PDR optimisation for the system with CMMSE channel estimator

MSE of the optimally designed CMMSE estimator has been given by (3.40) as

$$\sigma_{\varepsilon}^2 = P \,\mathrm{tr}[\mathbf{S}\, \mathbf{U}_{\underline{\mathbf{hh}}} \mathbf{\Lambda}_{\underline{\mathbf{hh}}} (P\, SNR_{\mathrm{p}} \mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1} \mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}]\,, \tag{4.18}$$

where $\mathbf{S} = \mathbf{S}(0)$.

Substituting SNR_{d} (4.10), SNR_{p} (4.11) and σ_{ε}^2 (4.18) into (4.8) yields

$$\begin{aligned}\gamma &= c_{\mathrm{H}} SNR \left(1 + \frac{P N_{\mathrm{tx}} SNR}{(P N^{-1} \kappa + b)} \mathrm{tr} \left[\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}} \mathbf{S}\, \mathbf{U}_{\underline{\mathbf{hh}}} \mathbf{\Lambda}_{\underline{\mathbf{hh}}} \left(\frac{P\, SNR\, \kappa}{(P N^{-1} \kappa + b)} \mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I} \right)^{-1} \right] \right) \frac{(P N^{-1} \kappa + b)}{SNR} \\ &= c_{\mathrm{H}} (P N^{-1} \kappa + b + P N_{\mathrm{tx}} SNR\, \mathrm{tr}[\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}} \mathbf{S}\, \mathbf{U}_{\underline{\mathbf{hh}}} \mathbf{\Lambda}_{\underline{\mathbf{hh}}} (P\, SNR\, \kappa / (P N^{-1} \kappa + b)\, \mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1}])\,, \end{aligned} \tag{4.19}$$

where matrices $\mathbf{\Lambda}_{\underline{\mathbf{hh}}}$ and $\mathbf{U}_{\underline{\mathbf{hh}}}$ are obtained from the eigenvalue decomposition $\mathbf{U}_{\underline{\mathbf{hh}}} \mathbf{\Lambda}_{\underline{\mathbf{hh}}} \mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}} = P^{-1} \mathbf{S}^{-1/2} \mathbf{R}_{\underline{\mathbf{hh}}} \mathbf{S}^{-1/2}$ (refer to Subsection 3.3.4.2 for details).

For the equispaced pilot pattern, $\mathbf{S} = P^{-1} \mathbf{I}$ and hence the NMSE expression (4.19) is reduced to

$$\gamma_{\text{es}} = c_{\text{H}}\left(P N^{-1}\kappa + b + N_{\text{tx}} SNR \sum_{l=0}^{L-1} \lambda_l \left(1 + \frac{\lambda_l P\, SNR\, \kappa}{P N^{-1}\kappa + b} \right)^{-1} \right)$$
$$= c_{\text{H}}\left(P N^{-1}\kappa + b + N_{\text{tx}} SNR \sum_{l=0}^{L-1} \frac{\lambda_l (P N^{-1}\kappa + b)}{P(\lambda_l SNR + N^{-1})\kappa + b} \right). \quad (4.20)$$

It can be shown (Appendix D.2) that γ is a convex function in κ ($\kappa > 0$). Hence it has a global minimum at $\kappa_{\text{opt}} = \arg\min_{\kappa>0} \gamma$, which can be found from the equation $\partial\gamma/\partial\kappa = 0$, where the derivative of γ with respect to the PDR is calculated (Appendix D.2) in the form of

$$\frac{\partial\gamma}{\partial\kappa} = c_{\text{H}} P N^{-1}\left(1 - P N_{\text{tx}} N\, SNR^2 b \,\text{tr}\left[\mathbf{U}_{\underline{\mathbf{hh}}}^{\text{H}} \mathbf{S}\, \mathbf{U}_{\underline{\mathbf{hh}}} \boldsymbol{\Lambda}_{\underline{\mathbf{hh}}}^2 \left(P(SNR\, \boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} + N^{-1}\mathbf{I})\kappa + b\mathbf{I}\right)^{-2}\right]\right). \quad (4.21)$$

For the equispaced pilot pattern, (4.21) is simplified to

$$\frac{\partial\gamma_{\text{es}}}{\partial\kappa} = c_{\text{H}} \frac{P}{N}\left(1 - N_{\text{tx}} N\, SNR^2 b \sum_{l=0}^{L-1} \frac{\lambda_l^2}{[P(\lambda_l SNR + N^{-1})\kappa + b]^2}\right). \quad (4.22)$$

Solution of the quadratic equation $\partial\gamma/\partial\kappa = 0$ does not exist in the closed-form for an arbitrary distribution of $\lambda_l, l \in [0, L-1]$. Hence one needs to apply a numerical search procedure to find a positive root, which would correspond to κ_{opt}. However, this approach is not practical for the unified and adaptive system design. Instead, here we try to find an approximate closed-form PDR expression for the general case of MMSE channel estimation.

4.2.3.1 Weak upper bound on the optimal PDR

To begin, we consider the special case of $\partial\gamma/\partial\kappa$ when $\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} = s P^{-1}\mathbf{I}$, where $s = 1/\text{tr}(\mathbf{S})$, that corresponds to $\mathbf{R}_{\underline{\mathbf{hh}}} = s\mathbf{S}$, leading to the worst-case performance of the channel estimator (refer to Subsection 3.3.4.2 for details). For the equispaced pilot arrangement, with $s = P/L$, this condition is interpreted as the case when all the principal components of the channel have the same average power, equal to $\lambda_l = L^{-1}, \forall l \in [0, L-1]$. Then γ (4.19) and $\partial\gamma/\partial\kappa$ (4.21) are correspondingly simplified to

$$\widetilde{\gamma} = c_{\text{H}}(P N^{-1}\kappa + b + N_{\text{tx}} SNR\, s\, \text{tr}[\mathbf{S}\,(SNR\, s\,\kappa/(P N^{-1}\kappa + b)\mathbf{I} + \mathbf{I})^{-1}])$$
$$= c_{\text{H}}(P N^{-1}\kappa + b)(1 + N_{\text{tx}} SNR/(a\kappa + b)) \quad (4.23)$$

and

$$\frac{\partial\widetilde{\gamma}}{\partial\kappa} = c_{\text{H}} P N^{-1}\left(1 - P^{-1} N_{\text{tx}} N\, SNR^2 b\, s^2 \left[P(P^{-1} SNR\, s + N^{-1})\kappa + b\right]^{-2} \text{tr}(\mathbf{S})\right)$$
$$= c_{\text{H}} P N^{-1}\left(1 - P^{-1} N_{\text{tx}} N\, SNR^2 b\, s\,(a\kappa + b)^{-2}\right)$$
$$= c_{\text{H}} P N^{-1} \frac{(a^2\kappa^2 + 2ab\kappa + b^2 - P^{-1} N_{\text{tx}} N\, SNR^2 b\, s)}{(a\kappa + b)^2}, \quad (4.24)$$

where $a = P N^{-1} + SNR\, s$.

Setting $\partial\widetilde{\gamma}/\partial\kappa = 0$, we arrive at the solution being a positive root of the quadratic equation

$$a^2\kappa^2 + 2ab\kappa + b^2 - P^{-1}N_{\mathrm{tx}}N\,SNR^2 bs = 0\,, \tag{4.25}$$

$$\begin{aligned}\widehat{\kappa}_{\mathrm{opt}} &= \frac{-ab + \sqrt{a^2b^2 - a^2(b^2 - P^{-1}N_{\mathrm{tx}}N\,SNR^2 bs)}}{a^2} \\ &= \frac{-b + SNR\sqrt{P^{-1}N_{\mathrm{tx}}N\,bs}}{a} \\ &= \frac{N\,SNR\sqrt{P^{-1}N_{\mathrm{tx}}N_{\mathrm{d}}s} - N_{\mathrm{d}}}{P + N\,SNR\,s}\end{aligned}\,, \tag{4.26}$$

where $N_{\mathrm{d}} = N - P\,N_{\mathrm{tx}}$ is the number of subcarriers within the block, which are designated to carry data symbols. For the high SNR values, one may approximate $\widehat{\kappa}_{\mathrm{opt}}$ as $\widehat{\kappa}_{\mathrm{opt}} \approx \sqrt{N_{\mathrm{tx}}N_{\mathrm{d}}/(s\,P)}$ that is known from (4.17) as the optimal PDR for the case of the CLS estimator.

For the equispaced pilot pattern, (4.26) takes the form of

$$\widehat{\kappa}_{\mathrm{opt,es}} = \frac{N\,SNR\sqrt{L^{-1}N_{\mathrm{tx}}N_{\mathrm{d}}} - N_{\mathrm{d}}}{P(1 + L^{-1}N\,SNR)} = \frac{N\,SNR\sqrt{L\,N_{\mathrm{tx}}N_{\mathrm{d}}} - L\,N_{\mathrm{d}}}{P(L + N\,SNR)}\,, \tag{4.27}$$

where the choice of $L = N_{\mathrm{cp}} + 1$ is made if the true order of the channel model is not known. According to the definition of the CIR correlation matrix (2.11), the true order of the channel model is equal to the number of multipaths K ($K \le N_{\mathrm{cp}} + 1$). However, in practice, there is no information about K in the system, hence a conservative assumption $L = N_{\mathrm{cp}} + 1$ is used to obtain (4.27).

Formulas (4.23) and (4.26) represent correspondingly detection NMSE and optimal PDR, peculiar to the largest CFR estimation MSE, $\widehat{\sigma}_{\varepsilon}^2$ (3.42), for a channel model of the anticipated order L. According to (4.8), where PDR, κ, is assumed to be fixed, the larger channel estimation MSE would also result in the greater detection NMSE γ. Hence hereafter we will term $\widehat{\gamma}$ (4.23) as the upper bound on γ. The corresponding PDR, $\widehat{\kappa}_{\mathrm{opt}}$ (4.26), will be referred to as a (*weak*) *upper bound* on κ_{opt}. The latter statement can be intuitively understood from the following argument. A more complicated (higher order) channel model incurs growth of the CFR estimation MSE. To compensate for that growth, one has to boost power of the pilot symbols, i.e. to increase PDR κ. The reasonable magnitude for PDR, κ_{opt}, is established as a trade-off between compensation of the channel estimation MSE and concurrent increase in the detection NMSE due to SNR lowering (proportional to κ^{-1}) on the data subcarriers. It is obvious that κ_{opt} is the highest for the most complicated channel model, i.e. having an order L and $\mathbf{R}_{\underline{\mathbf{hh}}} = s\mathbf{S}$.

In the subsequent analysis, a closed-form expression for the tight upper bound on κ_{opt} will be introduced specifically for the equispaced pilot arrangement. The relation between the weak and the tight PDR bounds will also be discussed.

4.2.3.2 Tight upper bound on the optimal PDR (equispaced pilot pattern only)

Consider the system with the equispaced pilot configuration, for which $\mathbf{S} = P^{-1}\mathbf{I}$. It has been shown in Appendix C.5 that the equispaced pilot pattern is optimal in the sense of minimising $\hat{\sigma}_{\varepsilon}^2$ (3.42) in comparison with other pilot arrangements.

In the light of the approximation introduced in (4.6), assuming a high SNR regime, it is of interest to analyse the dependence of PDR on SNR at the receiver input. Differentiating $\hat{\kappa}_{\text{opt,es}}$ (4.27) with respect to SNR yields

$$\begin{aligned}\frac{\partial \hat{\kappa}_{\text{opt,es}}}{\partial SNR} &= \frac{N\sqrt{L N_{\text{tx}} N_{\text{d}}}}{P(L + N\,SNR)} - \frac{N\,[N\,SNR\sqrt{L N_{\text{tx}} N_{\text{d}}} - L N_{\text{d}}]}{P(L + N\,SNR)^2} \\ &= N\sqrt{L N_{\text{d}}}\,\frac{[(L + N\,SNR)\sqrt{N_{\text{tx}}} - N\,SNR\sqrt{N_{\text{tx}}} + \sqrt{L N_{\text{d}}}\,]}{P(L + N\,SNR)^2} \\ &= N\,L\sqrt{N_{\text{d}}}\,\frac{[\sqrt{L N_{\text{tx}}} + \sqrt{N_{\text{d}}}\,]}{P(L + N\,SNR)^2}\end{aligned} \quad . \tag{4.28}$$

Since $\partial \hat{\kappa}_{\text{opt,es}} / \partial SNR > 0$ for any argument value, one can conclude that PDR is a monotonically increasing function of SNR.

Another important relationship is established by taking $\partial \hat{\kappa}_{\text{opt,es}} / \partial L$:

$$\begin{aligned}\frac{\partial \hat{\kappa}_{\text{opt,es}}}{\partial L} &= \frac{\frac{1}{2} N\,SNR\sqrt{\frac{N_{\text{tx}} N_{\text{d}}}{L}} - N_{\text{d}}}{P(L + N\,SNR)} - \frac{N\,SNR\sqrt{L N_{\text{tx}} N_{\text{d}}} - L N_{\text{d}}}{P(L + N\,SNR)^2} \\ &= \frac{\frac{1}{2} N\,SNR(L + N\,SNR)\sqrt{\frac{N_{\text{tx}} N_{\text{d}}}{L}} - N_{\text{d}}(L + N\,SNR) - N\,SNR\sqrt{L N_{\text{tx}} N_{\text{d}}} + L N_{\text{d}}}{P(L + N\,SNR)^2} \\ &= \frac{1/2\,N\,SNR(N\,SNR - L)\sqrt{N_{\text{tx}} N_{\text{d}}} - N N_{\text{d}} SNR\sqrt{L}}{P\sqrt{L}(L + N\,SNR)^2} \\ &= N\,SNR\sqrt{N_{\text{d}}}\,\frac{[1/2\,(N\,SNR - L)\sqrt{N_{\text{tx}}} - \sqrt{L N_{\text{d}}}\,]}{P\sqrt{L}(L + N\,SNR)^2}\end{aligned} \quad , \tag{4.29}$$

where $\partial \hat{\kappa}_{\text{opt,es}} / \partial L > 0$ if $1/2\,(N\,SNR - L)\sqrt{N_{\text{tx}}} - \sqrt{L N_{\text{d}}} > 0$, i.e. when $SNR > L/N + 2\sqrt{L N_{\text{d}} / (N^2 N_{\text{tx}})}$. The expression on the right side of the latter inequality, $L/N + 2\sqrt{L N_{\text{d}} / (N^2 N_{\text{tx}})}$, attains a maximum when $L/N = 1/4$ (the worst channel delay spread typically considered by the system design), $N_{\text{d}} = N$ (all subcarriers are used for data transfer), and $N_{\text{tx}} = 1$. Thus, one can evaluate this maximum in the form of the inequality $SNR > 5/4$. As $SNR >> 5/4$ in the normal operational conditions, it is proved that $\partial \hat{\kappa}_{\text{opt,es}} / \partial L > 0$, and hence $\hat{\kappa}_{\text{opt,es}}$ can be viewed as an increasing function of L.

In Subsection 3.3.4.2 it was shown that $\lambda_l = L^{-1}$, $\forall\, l \in [0, L-1]$ is known as the worst-case scenario in terms of the channel estimation MSE as it results in $\sigma_{\varepsilon,\text{es}}^2 = \hat{\sigma}_{\varepsilon,\text{es}}^2$ in (3.41). Hence $\kappa_{\text{opt,es}}$ is expected to be the highest for L equipowered principal components to minimise the adverse effect of inaccurate CFR estimation (later on this

intuitive statement will be illustrated by numerical examples). If the channel is known to have only K ($K \le L$) multipaths, then substitution of $L = K$ in (4.27) yields an upper bound on $\kappa_{\text{opt,es}}$, which is much tighter than $\widehat{\kappa}_{\text{opt,es}}$ with the conservative assumption of $L = N_{\text{cp}} + 1$. However, in practice, the true order of the channel model, K, is not known. Here we propose to determine the tight upper bound on $\kappa_{\text{opt,es}}$ on the basis of equation (4.27), where the *effective number of principal components* L_{eff} is used instead of the CIR length L. Assuming that the set of the eigenvalues of $\mathbf{R}_{\underline{\mathbf{hh}}}$, λ_l, is known or estimated at the receiver, L_{eff} can be computed from the equality

$$\sigma^2_{\varepsilon,\text{es}} = \sum_{l=0}^{L-1} \frac{\lambda_l}{1 + \lambda_l P \widehat{SNR}_{\text{p}}} = \frac{L_{\text{eff}}}{L_{\text{eff}} + P \widehat{SNR}_{\text{p}}}, \tag{4.30}$$

where the "guess" value of $\widehat{SNR}_{\text{p}}$ has to be used instead of the true SNR_{p} because the latter depends on PDR as per (4.11) and cannot be assessed without the knowledge of the optimum power allocation between pilot and data subcarriers.

Solving $\dfrac{L_{\text{eff}}}{L_{\text{eff}} + \widehat{\rho}} - \sum_{l=0}^{L-1} \dfrac{\lambda_l}{1 + \lambda_l \widehat{\rho}} = 0$, where $\widehat{\rho} = P \widehat{SNR}_{\text{p}}$, with respect to L_{eff} yields

$$L_{\text{eff}} = \widehat{\rho} \left[\left(\sum_{l=0}^{L-1} \frac{\lambda_l}{1 + \lambda_l \widehat{\rho}} \right)^{-1} - 1 \right]^{-1}. \tag{4.31}$$

Note that L_{eff} (4.31) can take non-integer values in general.

The dependence of L_{eff} on $\widehat{SNR}_{\text{p}}$ is revealed to be a non-decreasing function by observing that $\partial L_{\text{eff}}^{-1} / \partial \widehat{\rho} \le 0$. The proof of this is given in Appendix D.3. Hence one can conclude that specifying bigger guess value, $\widehat{SNR}_{\text{p}}$, creates an upper bound, L_{eff}, for the true order K of the channel model, that in turn causes overdetermining of PDR when substituted into (4.27) instead of L (this overdetermining is stipulated by the relationship $\partial \widehat{\kappa}_{\text{opt,es}} / \partial L > 0$).

A possible approach towards $\widehat{SNR}_{\text{p}}$ specification consists of two steps. First, the optimal PDR, $\widehat{\kappa}_{\text{opt,es}}$, is computed for the equipowered PDP as per (4.27). Second, $\widehat{SNR}_{\text{p}}$ is found based on $\widehat{\kappa}_{\text{opt,es}}$ according to

$$\widehat{SNR}_{\text{p}} = \frac{N \, SNR \, \widehat{\kappa}_{\text{opt,es}}}{P \, \widehat{\kappa}_{\text{opt,es}} + N_{\text{d}}}, \tag{4.32}$$

where SNR is recognised as the operational SNR at the receiver input.

Using the precomputed $\widehat{SNR}_{\text{p}}$ and L_{eff} values, the *tight upper bound* on $\kappa_{\text{opt,es}}$ is found as

$$\widehat{\widehat{\kappa}}_{\text{opt,es}} = \frac{N \, SNR \sqrt{L_{\text{eff}} N_{\text{tx}} N_{\text{d}}} - L_{\text{eff}} N_{\text{d}}}{P (L_{\text{eff}} + N \, SNR)}, \tag{4.33}$$

where $\widehat{\widehat{\kappa}}_{\text{opt,es}}$ attains minimum when $L_{\text{eff}} = K$.

4.2.4 PDR optimisation for the system with robust two-dimensional channel estimator

Derivation of the optimal PDR for the case of the 2D channel estimator is analogous to the cases of the intrablock estimators, considered in Subsections 4.2.2 and 4.2.3. Here we derive the closed-form expression of the upper bound on the optimal PDR for the case of the robust CLS-MMSE estimator based on the infinite Wiener filters. We show subsequently how to apply the resultant formula to the system with the finite-length filterbanks. We also introduce a hypothesis, based on the concept of the effective multipath model order developed in Subsection 4.2.3, that one can use the same formula with minor parameter alterations to obtain the upper bound on the optimal PDR for the CMMSE-MMSE channel estimator in the case of the equispaced pilot subcarriers.

4.2.4.1 Upper bound on the optimal PDR for the CLS-MMSE estimator with infinite memory

MSE of the robust CLS-MMSE estimator with $M = \infty$ has been given by (3.83) as

$$\sigma_\varepsilon^2 = \sum_{l=0}^{L-1} \rho_l \left[1 - \left(g_l \rho_l^{-1} \Omega + 1 \right)^{-\Omega^{-1}} \right], \tag{4.34}$$

where $g_l = [\mathbf{R}_{\underline{\underline{\mathbf{hh}}}}]_{l,l}$ (3.63), $\rho_l = SNR_\mathrm{p}^{-1} c_l$ (3.65) and $c_l = [\mathbf{S}(0)]_{l,l}$.

Substituting SNR_d (4.10), SNR_p (4.11) and σ_ε^2 (4.34) into (4.8) yields

$$\begin{aligned}
\gamma &= c_\mathrm{H} \left(P N^{-1} \kappa + b \right) \left(1 + \frac{N_\mathrm{tx}}{\kappa} \sum_{l=0}^{L-1} c_l \left[1 - \left(\frac{g_l}{c_l} \frac{\Omega \, SNR \, \kappa}{(P N^{-1} \kappa + b)} + 1 \right)^{-\Omega^{-1}} \right] \right) \\
&= c_\mathrm{H} \left[\frac{P}{N} \kappa + b + \frac{P N_\mathrm{tx}}{N} \sum_{l=0}^{L-1} c_l \left[1 - \left(\frac{(\Omega \, SNR \, g_l c_l^{-1} + P N^{-1}) \kappa + b}{P N^{-1} \kappa + b} \right)^{-\Omega^{-1}} \right] \right. \\
&\quad \left. + \frac{N_\mathrm{tx} b}{\kappa} \sum_{l=0}^{L-1} c_l \left[1 - \left(\frac{(\Omega \, SNR \, g_l c_l^{-1} + P N^{-1}) \kappa + b}{P N^{-1} \kappa + b} \right)^{-\Omega^{-1}} \right] \right] \\
&= c_\mathrm{H} \left[\frac{P}{N} \kappa + b + N_\mathrm{tx} \left(\frac{P}{N} + \frac{b}{\kappa} \right) \sum_{l=0}^{L-1} c_l \left[1 - \left(\frac{P \kappa + N b}{a_l \kappa + N b} \right)^{\Omega^{-1}} \right] \right],
\end{aligned} \tag{4.35}$$

where

$$a_l = \Omega N \, SNR \, g_l c_l^{-1} + P. \tag{4.36}$$

The function (4.35) is analytical with regard to the argument κ, in particular its first two derivatives exist in the closed form, as shown in Appendix D.4. As the second-order derivative $\partial^2 \gamma / \partial \kappa^2 > 0$ for $\kappa \in (0, \infty)$, the function (4.35) is convex in the interval of interest. Hence there is an optimal PDR $\kappa_\mathrm{opt} = \arg\min_{\kappa > 0} \gamma$, which can be found by solving

$$\frac{\partial \gamma}{\partial \kappa} = c_\mathrm{H} \left[\frac{P}{N} - \frac{N_\mathrm{tx} b}{s \kappa^2} + \frac{N_\mathrm{tx} b}{\kappa} \sum_{l=0}^{L-1} c_l \left(\frac{P \kappa + N b}{a_l \kappa + N b} \right)^{\Omega^{-1}} \left(\frac{1}{\kappa} + \frac{a_l - P}{\Omega (a_l \kappa + N b)} \right) \right] = 0. \tag{4.37}$$

The closed-form expression for the roots of (4.37) cannot be obtained; therefore one has to rely on numerical methods, which are not very convenient. A closer look at (4.37) reveals that the PDR-dependent term $\left((P\kappa+Nb)(a_l\kappa+Nb)^{-1}\right)^{\Omega^{-1}}$ has to be reduced, in order to bring the equation to the polynomial form. We propose to approximate it by a constant (later on the validity of this approximation will be explained)

$$\varphi_l=\left(\frac{P\kappa_{\text{eff}}+Nb}{a_l\kappa_{\text{eff}}+Nb}\right)^{\Omega^{-1}}, \tag{4.38}$$

so that (4.37) becomes

$$\frac{\partial\gamma}{\partial\kappa}\approx c_{\text{H}}\left[\frac{P}{N}-\frac{N_{\text{tx}}b}{s\kappa^2}+N_{\text{tx}}b\sum_{l=0}^{L-1}c_l\varphi_l\left(\frac{1}{\kappa^2}+\frac{a_l-P}{\Omega\kappa(a_l\kappa+Nb)}\right)\right]. \tag{4.39}$$

PDR κ_{eff}, selected for the approximation, is defined as the optimal PDR for the CLS estimator, which is seen as the primary module of the CLS-MMSE estimator. It is obvious that MSE of the CLS estimator is always higher than that of the CLS-MMSE estimator. Hence the CLS estimator is characterised by the higher optimal PDR magnitude. The latter is described by (4.17), being equal to

$$\kappa_{\text{eff}}=\sqrt{\frac{N_{\text{tx}}Nb}{sP}}. \tag{4.40}$$

By differentiating (4.39) with respect to κ, we obtain:

$$\begin{aligned}\frac{\partial^2\gamma}{\partial\kappa^2}&\approx c_{\text{H}}\left[\frac{2N_{\text{tx}}b}{s\kappa^3}-N_{\text{tx}}b\sum_{l=0}^{L-1}c_l\varphi_l\left(\frac{2}{\kappa^3}+\frac{(a_l-P)(2a_l\kappa+Nb)}{\Omega\kappa^2(a_l\kappa+Nb)^2}\right)\right]\\&=\frac{2N_{\text{tx}}b}{\kappa^3}\sum_{l=0}^{L-1}c_l\left[1-\varphi_l\left(1+\frac{(a_l-P)(a_l\kappa+Nb/2)}{\Omega(a_l+Nb\kappa^{-1})(a_l\kappa+Nb)}\right)\right].\end{aligned} \tag{4.41}$$

From here it is obvious that

$$\frac{\partial^2\gamma}{\partial\kappa^2}>\frac{2N_{\text{tx}}b}{\kappa^3}\sum_{l=0}^{L-1}c_l\left[1-\varphi_l\left(1+\frac{a_l-P}{\Omega(P+Nb\kappa_{\text{eff}}^{-1})}\right)\right], \tag{4.42}$$

since $a_l+Nb\kappa^{-1}>P+Nb\kappa_{\text{eff}}^{-1}$ if $\kappa\le\kappa_{\text{eff}}$, i.e. lies in the region of interest. Hence it follows from the expansion inequality (D.21), where κ is replaced by κ_{eff}, namely

$$\varphi_l^{-1}=\left(\frac{a_l\kappa_{\text{eff}}+Nb}{P\kappa_{\text{eff}}+Nb}\right)^{\Omega^{-1}}>1+\frac{1}{\Omega}\left(\frac{a_l-P}{P+Nb\kappa_{\text{eff}}^{-1}}\right), \tag{4.43}$$

that

$$\varphi_l\left(1+\frac{a_l-P}{\Omega(P+Nb\kappa_{\text{eff}}^{-1})}\right)<1 \tag{4.44}$$

and $\partial^2\gamma/\partial\kappa^2>0$. Thus, one can conclude that the approximated function (4.41) preserves behaviour of the true second-order derivative function (D.16). This observation stresses the weak influence of the term

$\left((P\kappa+Nb)(a_l\kappa+Nb)^{-1}\right)^{\Omega^{-1}}$ in the solution of the equation (4.37). Approximation, relying on κ_{eff}, guarantees that the function $\partial\gamma/\partial\kappa$ will be relatively precisely replicated in the zero region where the optimum PDR solution is found.

The upper bound on the detection NMSE is determined, substituting the upper bound on the channel estimation MSE given by (3.84), namely

$$\bar{\sigma}_{\varepsilon}^2 = \frac{1}{SNR_{\text{p}}}\left[\frac{1}{s} - \frac{L}{P}\left(\frac{\Omega P SNR_{\text{p}}}{L}+1\right)^{-\Omega^{-1}}\right], \tag{4.45}$$

into (4.8), i.e.

$$\bar{\gamma} = c_{\text{H}}\left[\frac{P}{N}\kappa + b + N_{\text{tx}}\left(\frac{P}{N}+\frac{b}{\kappa}\right)\left(\frac{1}{s} - \frac{L}{P}\left(\frac{P\kappa+Nb}{a\kappa+Nb}\right)^{\Omega^{-1}}\right)\right], \tag{4.46}$$

where

$$a = P(\Omega L^{-1} N SNR + 1). \tag{4.47}$$

Using the approximation (4.39), we express

$$\begin{aligned}\frac{\partial\bar{\gamma}}{\partial\kappa} &\approx c_{\text{H}}\left[\frac{P}{N} - \frac{N_{\text{tx}}b}{s\kappa^2} + \frac{LN_{\text{tx}}b\bar{\varphi}}{P\kappa}\left(\frac{1}{\kappa}+\frac{a-P}{\Omega(a\kappa+Nb)}\right)\right]\\ &= c_{\text{H}}\frac{PN^{-1}\kappa^2(a\kappa+Nb) - N_{\text{tx}}bs^{-1}(a\kappa+Nb) + P^{-1}LN_{\text{tx}}b\bar{\varphi}(a\kappa+Nb) + \Omega^{-1}P^{-1}LN_{\text{tx}}b\bar{\varphi}(a-P)\kappa}{\kappa^2(a\kappa+Nb)}\\ &= c_{\text{H}}\frac{PN^{-1}a\kappa^3 + Pb\kappa^2 + (-N_{\text{tx}}abs^{-1} + P^{-1}LN_{\text{tx}}ab\bar{\varphi} + N_{\text{tx}}N SNRb\bar{\varphi})\kappa - N_{\text{tx}}Nb^2s^{-1} + P^{-1}LN_{\text{tx}}Nb^2\bar{\varphi}}{\kappa^2(a\kappa+Nb)}\\ &= \frac{c_{\text{H}}N}{Pa}\frac{F(\kappa)}{\kappa^2(a\kappa+Nb)}\end{aligned}, \tag{4.48}$$

where

$$\bar{\varphi} = \left(\frac{P\kappa_{\text{eff}}+Nb}{a\kappa_{\text{eff}}+Nb}\right)^{\Omega^{-1}}, \tag{4.49}$$

$$F(\kappa) = \kappa^3 + A\kappa^2 + B\kappa + C \tag{4.50}$$

and

$$A = Nb/a, \tag{4.51}$$

$$B = \frac{N_{\text{tx}}Nb}{P}\left(\frac{N SNR\bar{\varphi}}{a} + \frac{L\bar{\varphi}}{P} - \frac{1}{s}\right), \tag{4.52}$$

$$C = \frac{N_{\text{tx}}N^2b^2}{Pa}\left(\frac{L\bar{\varphi}}{P} - \frac{1}{s}\right). \tag{4.53}$$

The upper bound on the optimal PDR, $\widehat{\kappa}_{\text{opt}}$, represents a solution of the cubic equation $F(\kappa)=0$. To identify this solution uniquely, one should characterise the function (4.50) analytically. This is done in Appendix D.5, leading to the resultant expression describing the upper bound on κ_{opt}:

$$\widehat{\kappa}_{\text{opt}} = x_1 - A/3 = \frac{2}{3}\sqrt{A^2-3B}\cos\left[\frac{1}{3}\arccos\left(\frac{9AB-27C-2A^3}{2\sqrt{(A^2-3B)^3}}\right)\right]-\frac{A}{3}. \tag{4.54}$$

As $A\to 0$ and contribution of the corresponding terms (including the coefficient C proportional to A) in (4.54) is negligible, approximating expression can be adopted instead of the exact formula, being substantially reduced to

$$\widehat{\kappa}_{\text{opt}} \approx \sqrt{-B} = \kappa_{\text{eff}}\sqrt{[1-s\,\widehat{\varphi}(P^{-1}L+N\,SNR\,a^{-1})]}, \tag{4.55}$$

which is found to be similar to κ_{eff} (4.40), except the reduction term, introduced due to the Wiener filtering.

4.2.4.2 MSE relation between Wiener filters with finite and infinite memory

The suboptimal PDR, described by (4.55), corresponds to the theoretical case when the Wiener filterbank module of the CLS-MMSE channel estimator is assumed to have infinite length ($M=\infty$). In real systems, complexity considerations impose limitations on the filter memory that leads to some degradation of the estimation accuracy (refer to Subsection 3.4.2 for details). Impossibility of the spectral performance analysis of the finite Wiener filter, in contrast to its infinite counterpart, challenges system optimisation as the pilot power allocation has to be reformulated for each individual *M* value, due to its impact on the channel estimation MSE. By the following numerical example we show that there exists a convenient MSE mapping from the finite to the infinite scenario that allows taking into account channel estimation MSE increase due to the filter length truncation, without altering the suboptimal PDR formula (4.55). Fig.4.1 plots NMSE of the robustly designed Wiener filter versus SNR at its input, g_l/ρ_l. It can be seen that the MSE increase because of the smaller *M* is equivalent to the MSE increase due to the higher Doppler shift $\tilde{\omega}_{\text{D}}$ (3.82), e.g., the curves, corresponding to $\tilde{\omega}_{\text{D}}=0.0025\pi$, $M=10$ and $\tilde{\omega}_{\text{D}}=0.025\pi$, $M=\infty$, demonstrate almost the same trend. Hence an elegant analytical interpretation follows: varying the filter length can be considered as virtual modulation of the Doppler shift. The "modulated" parameter $\overline{\Omega}$, replacing Ω (3.81) in (4.47), can be determined by numerically solving the integral equation

$$\frac{1}{\delta_{\max}-\delta_{\min}}\int_{\delta_{\min}}^{\delta_{\max}}\frac{\delta\left[1-\overline{\mathbf{r}}^{\text{H}}(\overline{\mathbf{R}}+\delta^{-1}\mathbf{I})^{-1}\overline{\mathbf{r}}\right]}{1-\left(\delta\overline{\Omega}+1\right)^{-\overline{\Omega}^{-1}}}\text{d}\delta=1, \tag{4.56}$$

where $\overline{\mathbf{r}}$ and $\overline{\mathbf{R}}$ are defined in Subsection 3.4.2.2, $\delta=g_l/\rho_l$, $\delta_{\min}\le\delta\le\delta_{\max}$, and selection of $\delta_{\min}$ and $\delta_{\max}$ depends on the operational SNR range, as well as the anticipated channel PDP, e.g., defined for the worst-case performance as described in Subsection 3.4.2.1.

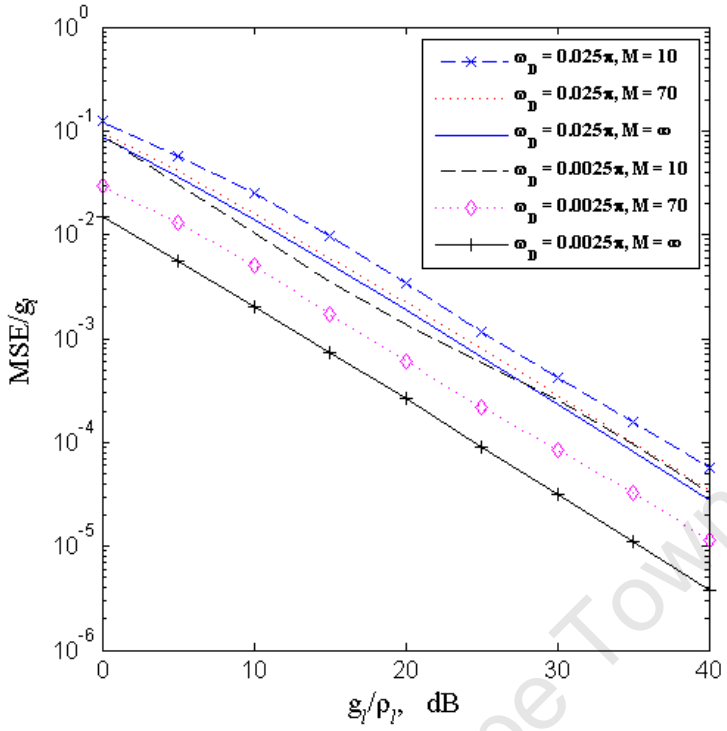


Fig.4.1. NMSE of the robust filter output as a function of g_l/ρ_l ($\omega_D = \tilde{\omega}_D = \overline{\omega}_D$)

4.2.4.3 Upper bound on the optimal PDR for the CMMSE-MMSE estimator (equispaced pilot pattern)

Taking into account the functional similarity between the intrablock smoothing modules of the CMMSE-MMSE estimator and the CMMSE estimator, we propose herein to compute the upper bound on the optimal PDR for the case of the CMMSE-MMSE estimator by specifying an effective order of the CIR model, similar to Subsection 4.2.3.2. This hypothetical approach is, however, limited to the equispaced pilot configurations. The effective order of the multipath model has been defined in (4.31) as

$$L_{\text{eff}} = P\,\widehat{SNR}_{\text{p}}\left[\left(\sum_{l=0}^{L-1}\frac{\lambda_l}{1+\lambda_l P\,\widehat{SNR}_{\text{p}}}\right)^{-1} - 1\right]^{-1}, \tag{4.57}$$

where λ_l are the eigenvalues of the intrablock CIR correlation matrix $\mathbf{R}_{\underline{\mathbf{hh}}}$, and $\widehat{SNR}_{\text{p}}$ is the predicted SNR at the pilot subcarriers, which can be expressed as

$$\widehat{SNR}_{\text{p}} = \frac{N\,SNR\,\hat{\kappa}_{\text{opt}}}{P\,\hat{\kappa}_{\text{opt}} + N - P\,N_{\text{tx}}}, \tag{4.58}$$

where $\hat{\kappa}_{\text{opt}}$ is the upper PDR bound (4.55), inherent to the CLS-MMSE estimator. The corresponding optimal PDR bound, $\hat{\tilde{\kappa}}_{\text{opt}}$, is then obtained by the replacement of L with L_{eff} in (4.55) (including the terms $s = P/L$, κ_{eff} (4.40), a (4.47) and $\hat{\varphi}$ (4.49)):

$$\widehat{\kappa}_{\text{opt,es}} \approx \kappa_{\text{eff}} \sqrt{\left[1-\left(\frac{P\kappa_{\text{eff}}+Nb}{a_{\text{eff}}\kappa_{\text{eff}}+Nb}\right)^{\Omega^{-1}}\left(1+PL_{\text{eff}}^{-1}N\,SNR\,a_{\text{eff}}^{-1}\right)\right]}, \tag{4.59}$$

where $a_{\text{eff}} = P(\Omega L_{\text{eff}}^{-1} N\,SNR+1)$ and $\kappa_{\text{eff}} = P^{-1}\sqrt{L_{\text{eff}} N_{\text{tx}} N b}$.

4.3 Numerical examples

Analytical PDR derivations, presented in Section 4.2 for the intrablock CMMSE and the 2D channel estimator scenarios, have been validated both by the direct substitution of numerical parameter values in the formulas of interest, i.e. (4.19), (4.21), (4.26), (4.31)-(4.33), (4.35), (4.55) and (4.57)-(4.59), and also with the help of the Monte Carlo simulation model. Subsection 4.3.1 lists most of the system parameters and outlines features of the model.

Subsections 4.3.2 and 4.3.3 deal with two case studies, which are specific to the adopted channel estimation algorithm; each is split into the numerical analysis part and the simulation validation part. The numerical analysis of the function, describing the upper bound on the optimal PDR, helps to understand the influence of various parameters (number of pilot subcarriers and Tx antennas, SNR, Doppler shift, multipath model order, etc.) on the optimal pilot settings. The main objective of the simulation results is to show that the optimal PDR (and its upper bound), determined based on the theoretical cost function, matches the actual optimal PDR, established by observing performance of the simulated system.

4.3.1 System configuration

Similar to the experiments in Sections 2.4 and 3.6, we consider a discrete-time pilot-assisted SM-MIMO-OFDM system with $N = 64$ subcarriers. Cyclic prefix (CP) length is set to $N_{\text{cp}} = 7$ to accommodate CIR with the modelled length of $L = 8$ samples. The number of pilot subcarriers per Tx antenna is set to either $P = 8$ (equispaced pilot pattern) or $P = 12$ (general pattern, with the pilot subcarrier arrangement in the band that is described in Section 3.1 as $p_k = k\lfloor N/P \rfloor,\ k \in [0, P-1]$). Pilot symbols are transmitted in each block (i.e. the pilot periodicity coefficient $\underline{P} = 1$). The SM configuration is tested in two modes: with 2 Tx and 3 Rx antennas and with 4 Tx and 6 or 10 Rx antennas. We consider only configurations with $N_{\text{rx}} > N_{\text{tx}}$, to avoid occurrence of the random cases when the transform matrix $\hat{\mathbf{H}}_{n,m}^{\text{H}}\hat{\mathbf{H}}_{n,m}$ in (4.5) is ill-conditioned and thus may negatively impact the detection statistics (refer to Subsection 2.3.3.1 for details).

The modelled channel is block-wise time-variant with the Doppler spectrum given by the Jakes model [118] and the maximum angular shift $\omega_{\text{D}} = 0.025\pi$, which represents the tolerable maximum according to (2.41). Three different multipath models have been considered:

Ch.1) $K = L$ sample-spaced independent identically distributed (IID) components with the flat power distribution, defined in (2.14) by the factor $\alpha = 0$ (flat PDP);

Ch.2) $K = 3$ equipowered non-sample-spaced IID components with the non-sample-spaced excess delays $\tau_0 = 0$, $\tau_1 = 1.7\,B^{-1}$ and $\tau_2 = 4.4\,B^{-1}$ (Fig.2.8), where B is the system bandwidth;

Ch.3) $K = L$ sample-spaced IID multipath components with the exponential power decay, defined in (2.14) by the factor $\alpha = 2$.

The modelled MIMO channel properties are in full accordance with the assumptions made in Subsection 2.3.1.

The receiver has been tested with the CMMSE, CLS-MMSE and CMMSE-MMSE channel estimators, with the latter two being robustly designed for $\overline{\omega}_{\mathrm{D}} = 0.025\pi$. The length of each filter in the Wiener filterbank has been chosen to be equal to $M = 70$ coefficients. According to (4.56), this corresponds to $\overline{\Omega} \approx 1/0.033$, modelling the MSE increase because of the finite filter length. Note that setting the FIR length $M = 70$ results in about 20% excess MSE at $g_l / \rho_l = 30\,\mathrm{dB}$ in comparison with the infinite filter.

We have used SNR gain at the output of the ZF detector with respect to the receiver input as a metric to compare performance for various PDR settings. This SNR gain is computed as

$$gain = \frac{\sigma_{\mathrm{d}}^2}{\mathrm{E}\left[\left\| \hat{\mathbf{x}}_{n,m} - \mathbf{x}_{n,m} \right\|^2 \right] SNR}, \tag{4.60}$$

where $\|.\|$ denotes the euclidean norm, and the vectors $\mathbf{x}_{n,m}$ and $\hat{\mathbf{x}}_{n,m}$ have been specified in Subsection 4.2.1. Expectation in the denominator of (4.60) has been replaced by averaging over all the simulated channel response realisations.

For *symbol error rate* (SER) performance evaluation, QPSK modulation and ZF-SQRD detector (refer to Subsection 2.3.3.3 for the implementation details) have been used in the simulation model.

For convenience of the reader the important system parameters are listed in Tab.4.2.

Tab.4.2. List of the main system parameters (parameters marked by the italic font represent the functions of the non-italic-marked parameters)

Parameter	Notation	Simulated value
Total number of subcarriers	N	64
CIR length	L	8
Number of pilot subcarriers	P	8, 12
Number of Tx antennas	N_{tx}	2, 4
SNR at receiver input	SNR	Varied
Theoretical detection NMSE	γ	Functions (4.19) and (4.35)
Optimal PDR	κ_{opt}	Solutions of (4.21) and (4.37)
Weak upper bound on optimal PDR (CMMSE case)	$\widehat{\kappa}_{\mathrm{opt}}$ (also κ_{wub})	Function (4.26)
Tight upper bound on optimal PDR (CMMSE case)	$\widehat{\widehat{\kappa}}_{\mathrm{opt}}$ (also κ_{tub})	Function (4.33)
Upper bound on optimal PDR (CLS/CMMSE-MMSE case)	$\widehat{\kappa}_{\mathrm{opt}}$ (also κ_{ub})	Functions (4.55) and (4.59)

4.3.2 System with CMMSE channel estimator

The first case study deals with PDR optimisation in the system with the optimally designed intrablock CMMSE estimator, which has been described for the MIMO application in Subsection 3.3.5.

4.3.2.1 Numerical parameter analysis

We first let the system have the equispaced pilot configuration ($P = 8$) and two transmit antennas. In the process of PDR derivations we have introduced several approximations in (4.6), assuming that the system will operate in the high SNR regime. Hence it is of interest to investigate how SNR affects the optimal PDR, and to assess implications of the approximations.

Fig.4.2 shows optimal PDR values, κ_{opt}, determined as a numerical solution of (4.21) for the given SNR at the receiver input. One can see that the function has an increasing nature and asymptotically tends to the fixed value, defined for $SNR \to \infty$ and the true CIR model order K (number of multipaths). However, the rapidity of the increase depends on the channel model. For the channels with the flat multipath intensity profile (Ch.1 and Ch.2), κ_{opt} quickly (already after 15dB) approaches the asymptote, whereas the exponential PDP model (Ch.3) implies a more gradual growth of κ_{opt} in the SNR region from 0dB to 60dB. Hence, in general, SNR information is needed to determine the optimum PDR accurately even if the system operates in the high SNR regime. κ_{opt} asymptote, defined for $SNR \to \infty$, can serve as a reliable PDR setting if no SNR information is available.

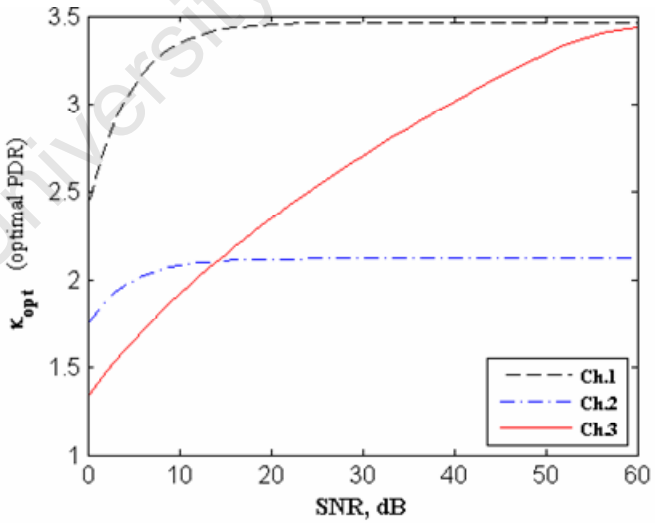


Fig.4.2. Optimal PDR, κ_{opt}, as a function of SNR ($P = 8$, $N_{\text{tx}} = 2$)

It is obvious for the case of the equispaced pilot placement that the increase of the number of pilot subcarriers leads to the decrease of the channel estimation MSE (4.18) that in turn yields κ_{opt} lowering according to the

arguments presented in Subsection 4.2.3.1. Note, however, that in general there exist some pilot patterns, which are not suitable for the channel estimator implementation. Such pilot patterns may cause a dramatic increase of the CFR estimation MSE (e.g., it is shown in Fig.3.11 for the system with CLS estimator, which is regarded as a primary module for the CMMSE estimator). Fig.4.3 shows κ_{opt} and $\widehat{\kappa}_{\text{opt}}$ (4.26) dependence on the number of pilot subcarriers, P, assuming their uniform arrangement when N/P is not an integer (as described in Subsection 4.3.1). The SNR is set to 30dB. It is significant to note that for the channel with L equipowered multipath components (Ch.1), κ_{opt} closely approaches its upper bound, $\widehat{\kappa}_{\text{opt}}$, while for the channel model of the smaller order (Ch.2) κ_{opt} lies far below $\widehat{\kappa}_{\text{opt}}$. The dependence of κ_{opt} on P is indeed decreasing for Ch.2. The same trend is observed for most P values in parts of the $\widehat{\kappa}_{\text{opt}}$ curve and κ_{opt} for Ch.1. A number of surges in the $\widehat{\kappa}_{\text{opt}}$ function correspond to the ill-conditioned estimator design modes (these surges are less noticeable in the Ch.2 graph). One should avoid using these modes. Alternatively, one can try to find a different pilot placement (e.g., as proposed in the work [91]), which would make well-conditioned estimator implementation feasible. Omitting the surge sections of $\widehat{\kappa}_{\text{opt}}$ in Fig.4.3, it is revealed that the difference between the optimum PDR (Ch.2) and its weak upper bound diminishes as the number of pilots grows.

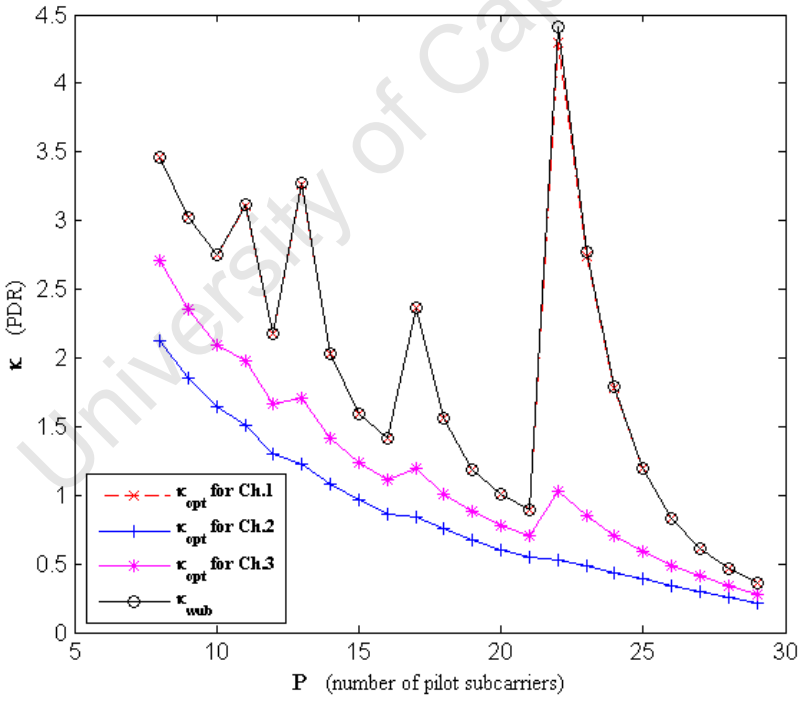


Fig.4.3. Optimal PDR, κ_{opt}, and weak upper bound on PDR, κ_{wub}, as a function of the number of pilot subcarriers ($N_{\text{tx}} = 2$, $SNR = 30\,\text{dB}$)

Fig.4.4 describes dependence of $\hat{\kappa}_{\text{opt}}$ on the two parameters: number of pilots, P, and the number of Tx antennas, N_{tx}, under the constraint (3.4), $L \le P \le N/N_{\text{tx}}$. Here the modes, corresponding to the non-decreasing sections of the $\hat{\kappa}_{\text{opt}}$ function in Fig.4.3, have been excluded as they do not satisfy the condition of performance improvement under the increase of the number of pilot subcarriers. It is interesting to point out that $\hat{\kappa}_{\text{opt}}$ reaches maximum at $N_{\text{tx}} = 4$ and $P = 8$, diagonally shifting down the plot to $N_{\text{tx}} = 1$ and $P = 31$.

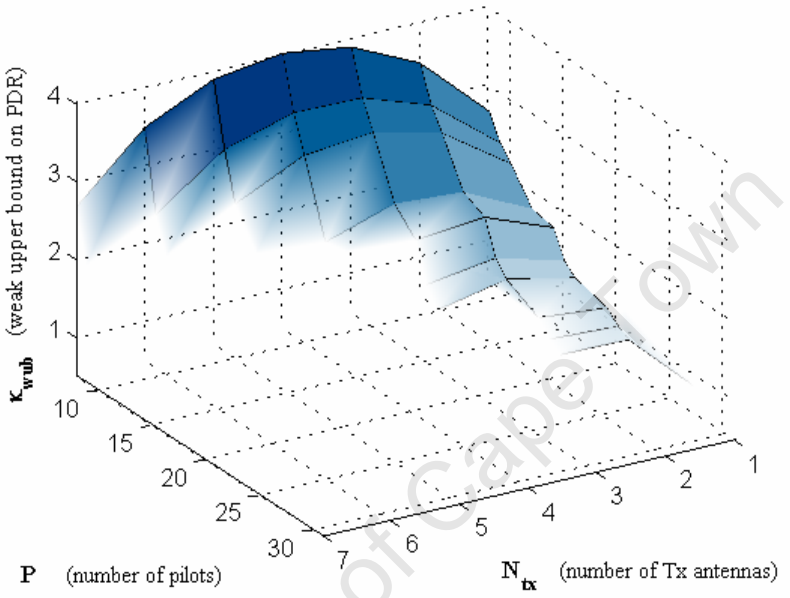


Fig.4.4. Weak upper bound on PDR, κ_{wub}, as a function of the number of pilot subcarriers and the number of Tx antennas ($SNR = 30\,\text{dB}$)

In Fig.4.5 and Fig.4.6 the parametric relationships, associated with the tight upper bound on PDR, $\hat{\bar{\kappa}}_{\text{opt,es}}$ (4.33), are shown. The corresponding system is characterised by $P = 8$ (equispaced pilot pattern), $N_{\text{tx}} = 2$ and the channel model Ch.3, where the exponential factor, α, is varied. The numerical illustration in Fig.4.5 confirms that L_{eff} is a non-decreasing function of $\hat{SNR}_{\text{p}}$ in (4.31), justifying selection of bigger $\hat{SNR}_{\text{p}}$ values for L_{eff} and hence calculation of $\hat{\bar{\kappa}}_{\text{opt,es}} \ge \kappa_{\text{opt,es}}$. $\hat{\bar{\kappa}}_{\text{opt,es}}$ values, along with $\kappa_{\text{opt,es}}$ and $\hat{\kappa}_{\text{opt,es}}$, are plotted in Fig.4.6 for a known range of SNRs. One can see that the model, characterised by the fastest PDP decay ($\alpha = 5$) and hence lesser effective number of the principal components, L_{eff}, implies a remarkable difference between the optimum PDR, $\kappa_{\text{opt,es}}$, and the weak upper bound on PDR, $\hat{\kappa}_{\text{opt,es}}$. In contrast to $\hat{\kappa}_{\text{opt,es}}$, the difference between $\hat{\bar{\kappa}}_{\text{opt,es}}$ and $\kappa_{\text{opt,es}}$ always remains small. This fact serves a significant motivation to estimate the CIR correlation matrix at the receiver and to determine its eigenvalues (principal components), which can be used afterwards to compute $\hat{\bar{\kappa}}_{\text{opt,es}}$.

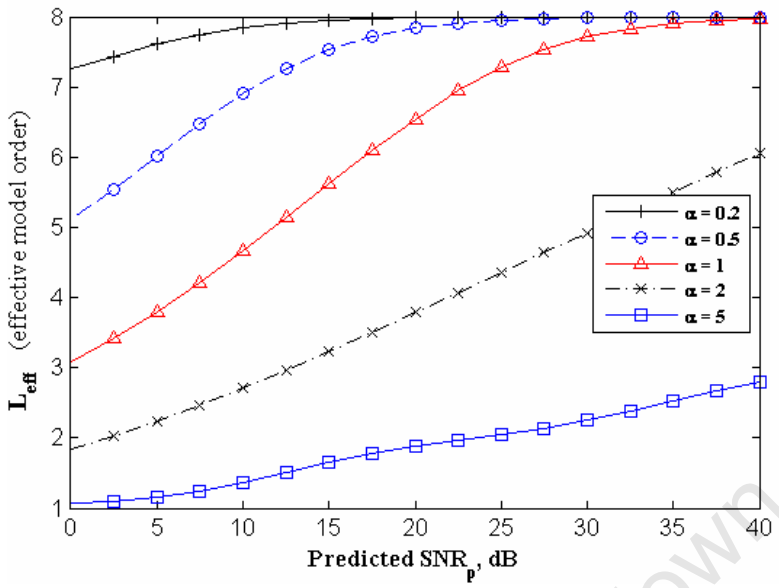


Fig.4.5. Dependence of the effective model order of the channel (Ch.3) on the predicted SNR at the pilot subcarriers ($P = 8$)

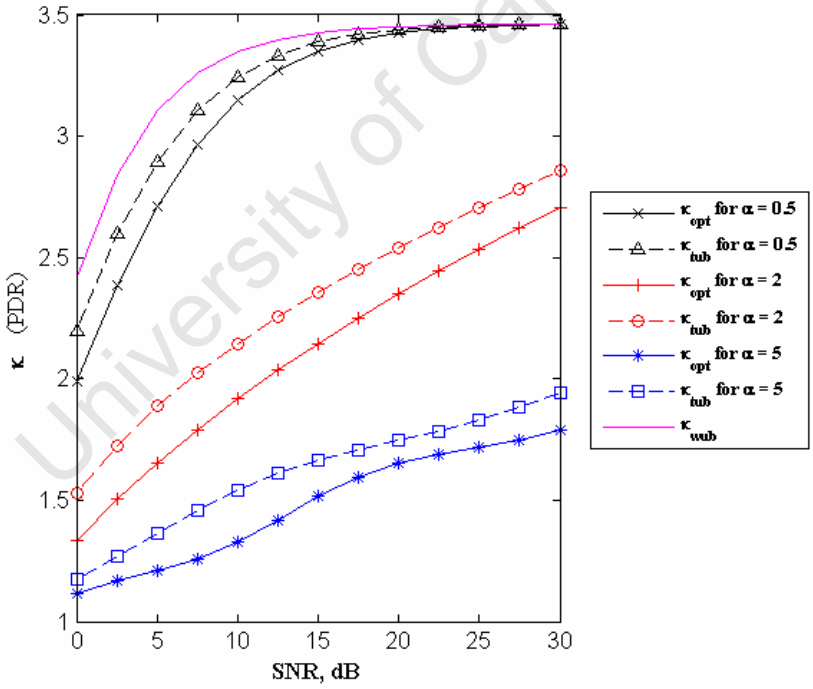


Fig.4.6. Optimal PDR value, κ_{opt}, tight upper bound, κ_{tub}, and weak upper bound on PDR, κ_{wub}, for different distributions of principal components (Ch.3) and SNR regimes ($P = 8$, $N_{tx} = 2$)

4.3.2.2 Simulation results

Fig.4.7 and Fig.4.8 show the dependence of SNR gain at the output of the detector on PDR for the three channel models when SNR at the receiver input is equal to 15dB or 30dB. The figures include both the simulated SNR gain results (4.60) and the theoretical SNR gain computed according to (4.19), where the factor c_{H} is determined by means of averaging over all the simulated channel response realisations. Optimal PDR, as well as weak and tight upper bounds are also shown in Fig.4.7 and Fig.4.8 for most of the specified system and channel configurations.

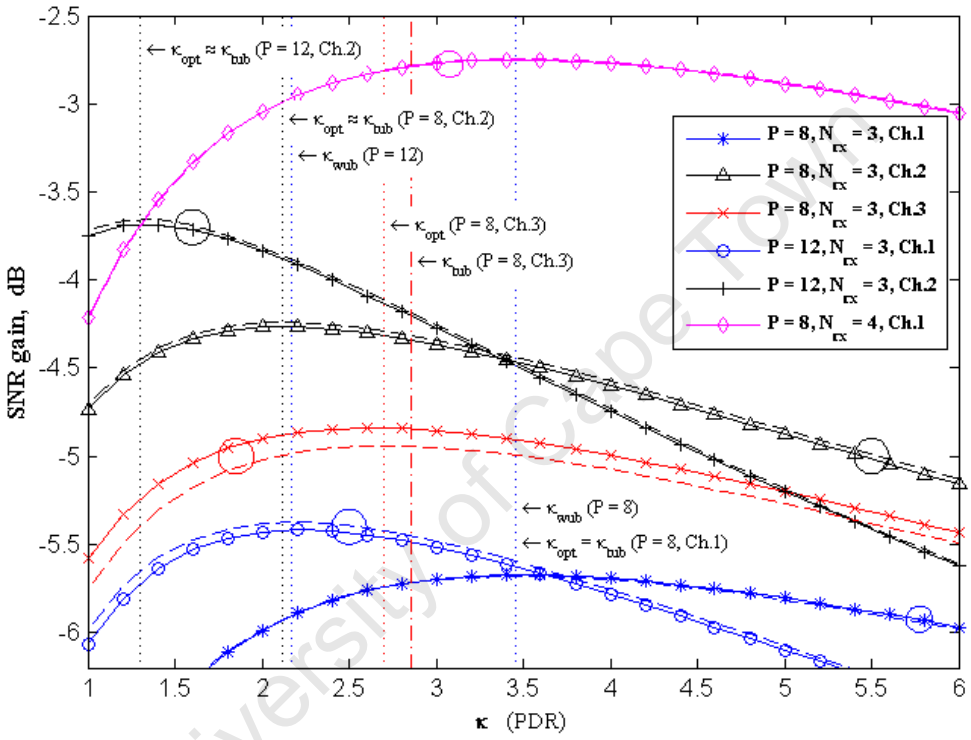


Fig.4.7. SNR gain of the detector output for the case of $N_{\mathrm{tx}} = 2$ (dashed-line theoretical curves, corresponding to the marker-labeled simulation curves, are indicated by ellipses)

The common feature of all curves is a flat grade at higher PDRs, i.e. at $\kappa > \kappa_{\mathrm{opt}}$. This observation justifies overdetermining of PDR when the exact operational conditions (SNR and the principal components of the channel) are not known. One can also see that the theoretical and the simulation SNR gain curves match each other in curvature, though the scaling might be quite distinct, and it is bigger for the lower SNRs (e.g., see the graphs corresponding to Ch.1 and Ch.3 in Fig.4.8). Curvature match is very important in the sense that the location of maximum is not changed, despite the approximations introduced in (4.6) and (4.7).

Fig.4.7 shows that the increase in the number of pilots leads to the shift of κ_{opt} in the negative direction (as already pointed out in Fig.4.3 and Fig.4.4) and improvement of the SNR gain. For the channels with a flat multipath

intensity profile (Ch.1 and Ch.2), $\tilde{\tilde{\kappa}}_{\text{opt}}$ has been found to be approximately equal to κ_{opt}. Despite the fact that $\tilde{\tilde{\kappa}}_{\text{opt}}$ has been derived only for the equispaced pilot configuration, one still can try to approximate κ_{opt} for a general pilot pattern using formulas (4.31)-(4.33), where $\tilde{\kappa}_{\text{opt}}$ is used instead of $\tilde{\kappa}_{\text{opt,es}}$. In such a case $\tilde{\tilde{\kappa}}_{\text{opt}}$ loses the interpretation of the tight upper bound, but represents just an approximation of κ_{opt}. The numerical example of $\tilde{\tilde{\kappa}}_{\text{opt}}$ computed for the 12 uniformly spaced pilots (Fig.4.7) confirms the accuracy of this approximation. For the channel model Ch.3, the difference between $\tilde{\tilde{\kappa}}_{\text{opt}}$ and κ_{opt} is more noticeable than for Ch.1 and Ch.2. Nonetheless, $\tilde{\tilde{\kappa}}_{\text{opt}}$ still lies in the quasi-optimal region, unlike $\tilde{\kappa}_{\text{opt}}$, and thus spares numerical solving of (4.21) to find an approximate value of the true optimal PDR. SNR gain difference, observed for the Ch.2 case between the $\tilde{\tilde{\kappa}}_{\text{opt}}$ and $\tilde{\kappa}_{\text{opt}}$ settings, is found to be about 0.2dB.

Fig.4.8 repeats the trends described in the context of Fig.4.7. Here the SNR gain difference for the Ch.2 case between $\tilde{\tilde{\kappa}}_{\text{opt}}$ and $\tilde{\kappa}_{\text{opt}}$ is approximately 0.25dB. With respect to different SNR regimes, it should be noted that $\tilde{\tilde{\kappa}}_{\text{opt}}$ attains a closer value to κ_{opt} if SNR is higher (Ch.3 graphs in Fig.4.8).

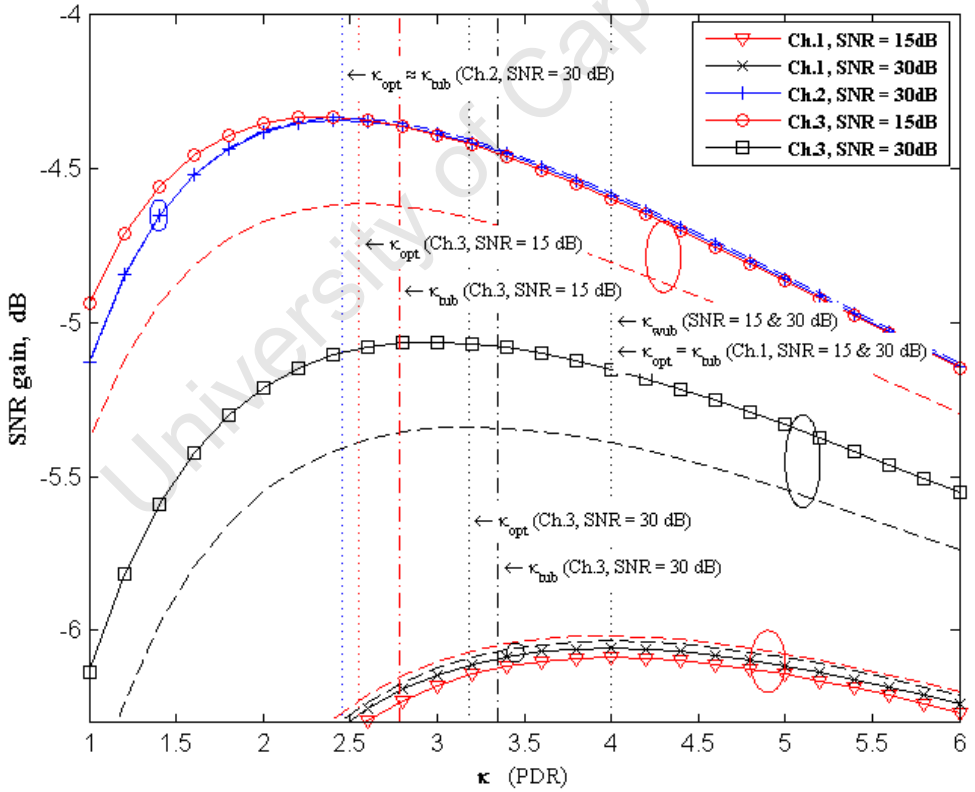


Fig.4.8. SNR gain of the detector output for the case of $N_{\text{tx}} = 4$, $N_{\text{rx}} = 6$, $P = 8$ (dashed-line theoretical curves, corresponding to the marker-labeled simulation curves, are indicated by ellipses)

Both the analysis (Subsection 4.2.3) and simulation results (coincidence of the corresponding maximums of $N_{rx}=3$ and $N_{rx}=4$ in Fig.4.7) establish that the optimal PDR does not depend on the order of the receive diversity. However, for the increased number of Rx antennas, dependence of the SNR gain on PDR may become stronger, necessitating selection of κ close to the optimal. This trend is observed in Fig.4.9, where for $N_{rx}=10$ there is up to 0.3dB SER decrease with the optimal PDR / tight upper bound, $\widehat{\kappa}_{opt,es}$, compared to the weak upper bound, $\widehat{\kappa}_{opt,es}$, whereas for $N_{rx}=6$ it does not exceed 0.1dB. Furthermore, this difference tends to grow with SNR.

Finally, one can see from Fig.4.9 the importance of proper PDR selection. If PDR is underdetermined, e.g., when pilot and data symbols have the same power in the considered setup, remarkable performance loss is inevitable. At the same time, specifying PDR from the flat grade region (like $\widehat{\kappa}_{opt}$) does not have such a negative impact on SER.

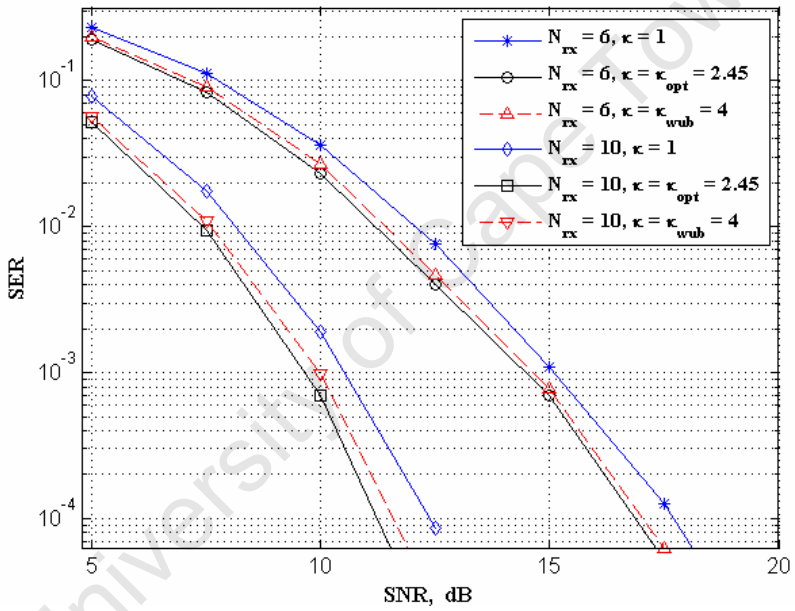


Fig.4.9. SER performance of the system with $P=8$, $N_{tx}=4$ and various PDR settings specified for the channel model Ch.2 and $SNR=30$dB

4.3.3 System with 2D channel estimator

It has been shown in Subsection 3.6.2 that application of the 2D estimation algorithms can sometimes yield a significant improvement of the channel estimates' accuracy in comparison with the intrablock processing methods. Therefore one could expect a shift in the optimal PDR to a lower magnitude, due to the reduction of the channel estimation MSE (and hence detection MSE) as a result of the better channel response filtering at the pilot

subcarriers. The second case study aims to illustrate this trend. Here we consider two 2D channel estimator types: CLS-MMSE and CMMSE-MMSE, which have been described for the MIMO application in Subsection 3.4.7.

4.3.3.1 Numerical parameter analysis

Consider first the optimal PDR for the system with 4 Tx antennas, 8 equispaced pilot subcarriers and the infinite CLS-MMSE estimator. This optimal PDR is determined by the numerical solution of (4.37), where we let $\Omega = 1/0.025$ as an example. From Fig.4.10 it becomes clear that κ_{opt} is an increasing function of SNR at the receiver input (this has also been observed earlier in Fig.4.2). One can also see that for Ch.3 κ_{opt} is lesser than for Ch.2. This could be explained by a faster Ch.3 PDP decay and hence smaller number of CIR samples, which require accurate estimation. The upper bound on the optimal PDR, $\tilde{\kappa}_{\text{opt}}$, is expected to have the same trend with regard to SNR as it can be regarded as κ_{opt} specified for the flat PDP ($g_l = L^{-1}$, $\forall l \in [0, L-1]$).

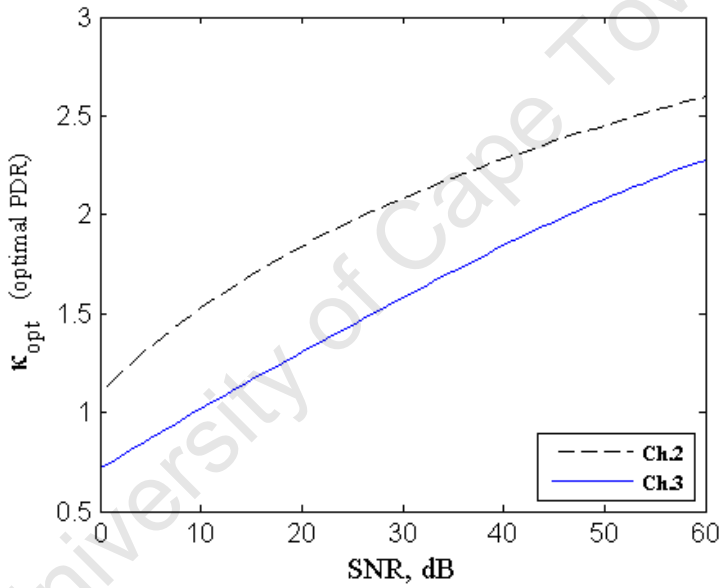


Fig.4.10. Optimal PDR, κ_{opt}, as a function of SNR ($P = 8$, $N_{\text{tx}} = 4$)

Fig.4.11 illustrates impact of the channel parameters on the upper PDR bound, $\tilde{\kappa}_{\text{opt}}$ (4.55), for the same system configuration. Here the integer CIR length L is replaced by the effective channel model order L_{eff}, which is a non-integer in general, as defined in Subsection 4.2.4.3. It can be seen that $\tilde{\kappa}_{\text{opt}}$ increases with both L_{eff} and Ω^{-1} ($\tilde{\omega}_{\text{D}}$). The variation of $\tilde{\kappa}_{\text{opt}}$ in L_{eff} is quite remarkable, necessitating analysis of the intrablock CIR statistics to raise system performance to the optimum level.

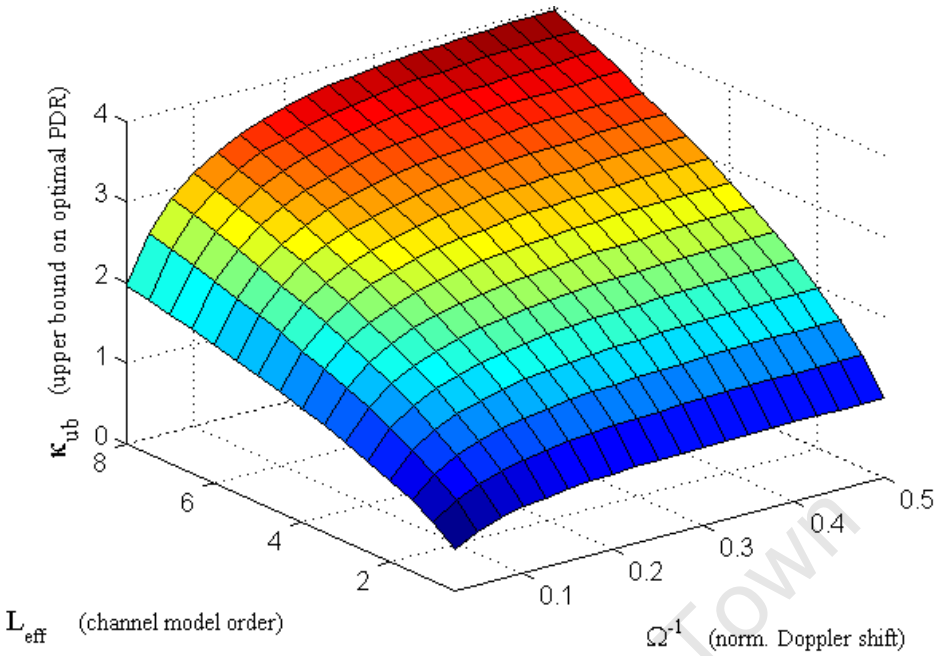


Fig.4.11. Dependence of upper PDR bound, κ_{ub}, on the effective channel model order, L_{eff}, and the normalised Doppler shift, Ω^{-1} (for the equispaced pilot pattern and $SNR = 30\,\text{dB}$)

4.3.3.2 Simulation results

Fig.4.12, Fig.4.13 and Fig.4.14 show the dependence of SNR gain at the detector output on PDR for the two channel models (Ch.2 and Ch.3). The figures include both the simulated SNR gain results (4.60) and the theoretical SNR gain, computed for the CLS-MMSE channel estimator as the inverse of (4.35), where the factor c_H is determined by means of averaging over all the simulated channel response realisations. Optimal PDR and its upper bound are also shown in the graphs for most of the specified system and channel configurations.

The common feature of all curves is a flat grade at higher PDRs, i.e. at $\kappa > \kappa_{opt}$. This observation justifies overdetermining of PDR, when the optimal PDR cannot be computed.

The theoretical and the simulation SNR gain curves match each other in curvature (Fig.4.12 and Fig.4.13), eliminating any uncertainty in location of the extremum region, though the scaling of the curves might be quite distinct. The difference between the theoretical and the simulation graphs is bigger for the lower SNRs, as can be seen from Fig.4.13. This could be explained by the high-SNR assumptions made to derive (4.8) in Subsection 4.2.1.

For the system with the CLS-MMSE channel estimator, the upper bound on the optimal PDR, $\hat{\kappa}_{opt}$, serves as a much better (closer to the optimum) PDR setting than κ_{eff} (4.40), which corresponds to the system with the intrablock CLS estimator, and which is also known from Subsection 4.2.3.1 as the weak upper bound on the optimal PDR in case of the intrablock CMMSE estimator and $SNR \to \infty$. This is clearly visible in Fig.4.12 for both the equispaced ($P = 8$) and the non-equispaced ($P = 12$) pilot pattern. If all the channel PDP samples are different from

zero (like in Ch.2), $\widehat{\kappa}_{\text{opt}}$ and κ_{opt} almost coincide. Note that, according to Fig.4.12, for the case of Ch.3, there is about 0.5dB performance gain when setting PDR to $\widehat{\kappa}_{\text{opt}}$ rather than κ_{eff}, whereas for the case of Ch.2, it is 0.3dB for both pilot patterns. At the same time the performance loss between $\widehat{\kappa}_{\text{opt}}$ and κ_{opt} is no more than 0.1dB for Ch.3, whereas for Ch.2 it is negligibly small as $\widehat{\kappa}_{\text{opt}}$ lies in the quasi-optimal region.

Fig.4.13 repeats the trends illustrated in Fig.4.12. For the lower SNR value (15dB) and the higher transmit and receive diversity, the difference between $\widehat{\kappa}_{\text{opt}}$ and κ_{eff} becomes even bigger: about 1dB for Ch.3 and 0.8dB for Ch.2. It should be pointed out here that unlike κ_{eff}, κ_{opt} and hence $\widehat{\kappa}_{\text{opt}}$ tend to grow with SNR. It is also shown in Fig.4.10. Thus, it is desirable to take the SNR information into account when setting PDR.

Simulations results plotted in Fig.4.14 demonstrate that the CMMSE-MMSE estimator improves performance of the system in contrast to the CLS-MMSE estimator only when the channel is characterised by the intrablock CIR correlation, i.e. when $\mathbf{R}_{\underline{\mathbf{hh}}}$ is non-diagonal (only in case of Ch.2). According to Ch.2 graphs, the performance of the CMMSE-MMSE configuration is about 0.9dB higher than that of CLS-MMSE, assuming optimal PDR settings. The upper bound on the optimal PDR in case of the CMMSE-MMSE-driven system has been found well-predicted based on the formulas (4.57) and (4.58) for both Ch.2 and Ch.3.

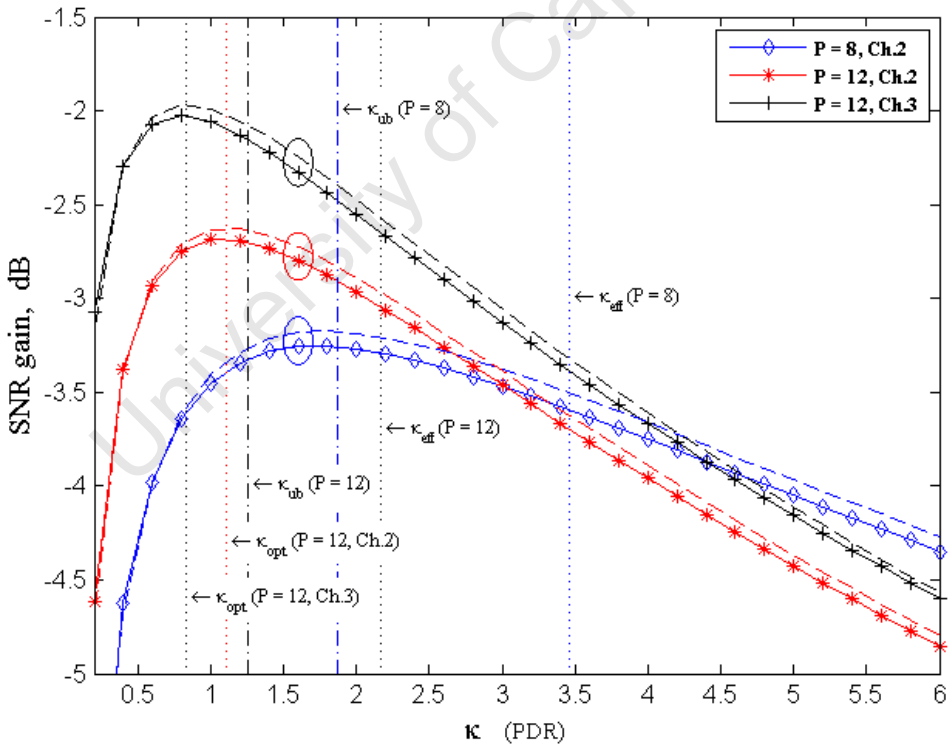


Fig.4.12. SNR gain of the detector output for the case of $N_{\text{tx}} = 2$, $SNR = 30\,\text{dB}$ and CLS-MMSE channel estimator (dashed-line theoretical curves, corresponding to the marker-labeled simulation curves, are indicated by ellipses)

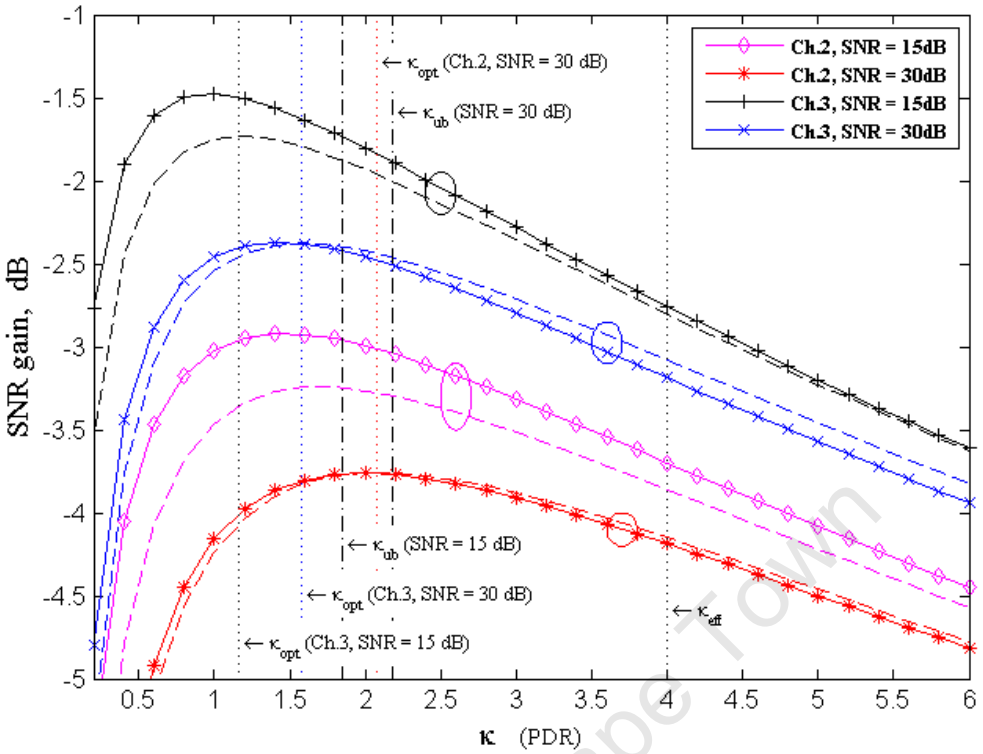


Fig.4.13. SNR gain of the detector output for the case of $N_{tx} = 4$, $P = 8$ and CLS-MMSE channel estimator (dashed-line theoretical curves, corresponding to the marker-labeled simulation curves, are indicated by ellipses)

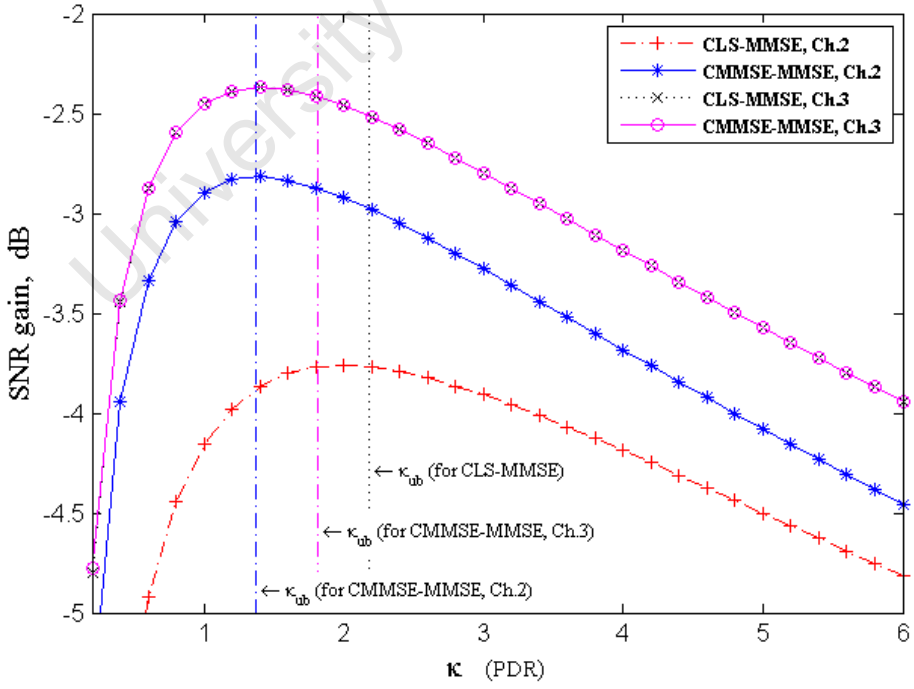


Fig.4.14. SNR gain of the detector output for $N_{tx} = 4$, $P = 8$, $SNR = 30\,\text{dB}$ and different channel estimators

In terms of the SER performance (Fig.4.15), the system with $\hat{\kappa}_{\text{opt}}$ PDR setting is up to 0.9dB better than the system having PDR set to κ_{eff} . This conclusion matches well the analysis of Fig.4.13. According to Fig.4.15, the CMMSE-MMSE configuration outperforms CLS-MMSE by 0.6dB-0.7dB if PDR is set to $\hat{\kappa}_{\text{opt}}$ corresponding to CMMSE-MMSE, and by 0.5dB if PDR is set to $\hat{\kappa}_{\text{opt}}$ corresponding to CLS-MMSE. The latter difference (0.1dB-0.2dB) is not very substantial and indicates that $\hat{\kappa}_{\text{opt}}$ (4.55) serves as a robust PDR setting for both the system with and without the smoothing module in the channel estimator.

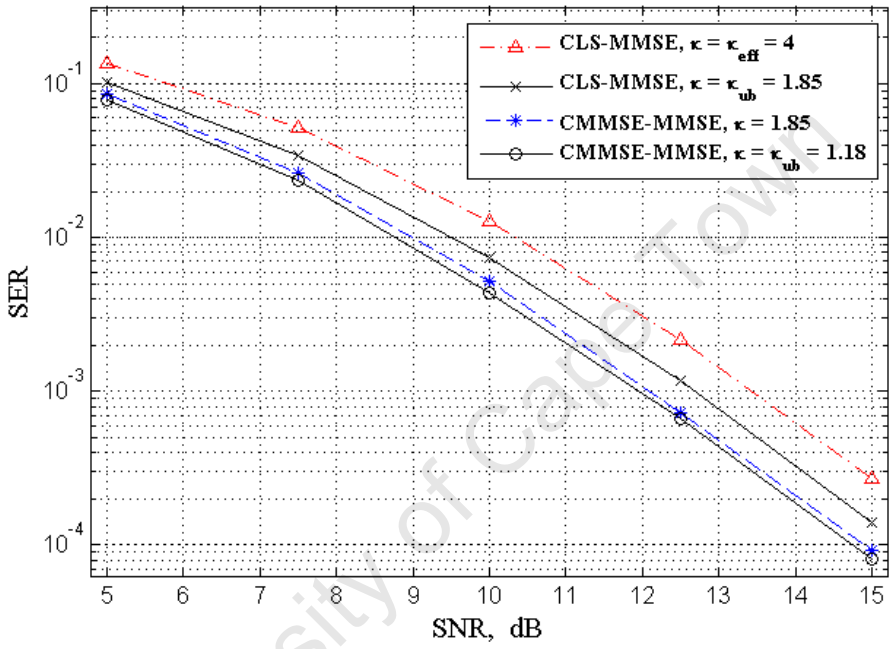


Fig.4.15. SER performance of the system with $N_{\text{tx}} = 4$, $N_{\text{rx}} = 6$, $P = 8$ and various PDR settings specified for the channel model Ch.2 and $SNR = 30\,\text{dB}$

4.4 Summary

When considering a composite system design, selection of the most suitable channel estimation or data detection algorithm is not sufficient by itself to guarantee the best performance of the system. As different portions of the transmitted signal (e.g., pilot and data symbols) are associated with different processing functions at the receiver, the structure of the signal should be optimised to minimise the error rate as seen at the demodulator output, and hence achieve the highest transmission capacity for the given class of the transceiver components characterised by the certain complexity level. Such a system optimisation represents a multivariate task, which can have partial

solutions at the specific layers of the system architecture. In this chapter, we have covered the two parts of the problem with application to the pilot-assisted SM-MIMO-OFDM framework: bandwidth and power allocation between data and pilot symbols. We have emphasised optimisation of the pilot-to-data power ratio, with the objective to minimise the variance of error at the detector output. It has been shown by several numerical case studies that PDR is a very important transmitter design parameter, which is regarded as a function of a variety of other parameters, namely the total number of subcarriers, number of pilot subcarriers, number of transmit antennas, multipath and Doppler model orders, SNR, etc. As a main contribution, we have derived the closed-form analytical expressions for the upper bound on the optimal PDR for the system with several channel estimator types. The upper bound on the optimal PDR can be viewed as a suboptimal PDR setting (quasi-optimal approximation), which has only a minor difference with the true optimum in the system performance sense. The obtained equations are applicable to the system with an arbitrary arrangement of the equipowered pilot subcarriers in the frequency band. The latter condition is very important, as a number of the hardware design constraints (e.g., virtual subcarriers introduced with the spectrum shaping purpose, parts of the band reserved for other channels or systems, etc.) might hamper the use of the optimal equispaced pilot pattern in practice.

It should be noted that analytical PDR expressions (even approximate) are much more preferable to the numerical methods of computing the optimal power allocation between the pilot and the data symbols due to the simplicity of their application. This fact favours applicability of the derived results to adaptive systems, which are capable of adjusting transmit signal configuration (e.g., block length, number of pilot subcarriers or antennas) according to the established channel conditions. In such systems, an optimal parameter tuning is desirable for each operational mode. The relatively simple closed-form PDR expressions, presented in this chapter, are designated to alleviate the challenging task of the on-the-fly optimisation of the multi-mode adaptive SM-MIMO-OFDM system.

5 Conclusions and Future Work

In this thesis, a pilot-assisted MIMO-OFDM receiver architecture, incorporating channel estimation and data symbol detection, has been investigated. We have considered a variety of algorithms and analysed their applicability in a number of channel environments. Furthermore, to achieve the maximum symbol error rate performance, the pilot structure has been optimised.

This chapter presents a summary of the most important findings, and the major conclusions deduced from the analyses in Chapters 2-4. This is followed by recommendations on possible system extensions to yield higher transmission capacity and lower symbol error rate. We also discuss the problem of optimisation of pilot-assisted MIMO-OFDM transmissions in more complex channel environments, which the next generation wireless communication systems will likely have to cope with.

5.1 General conclusions

In the SM-MIMO-OFDM systems, transmission capacity increase is achieved by adding more spatially multiplexed streams at the transmitter side, i.e. utilising more Tx antennas. Subject to the proper geometric design of the transceiver (spacing of the antenna elements), channel response at different spatial layers (SLs) can be regarded statistically independent. Hence in the absence of noise, an optimum receiver can incorporate a simple linear detector based on zero forcing (ZF) criteria. Additive noise tends to deteriorate linear ZF detection performance. This effect is stronger pronounced in the systems with minimum receive diversity. Replacing ZF with the MMSE linear detection scheme prevents dramatic enhancement of the detection noise. By capitalising on the property that detection represents a problem of the finite-alphabet solution search, linear detectors can be extended to decision-feedback structures. It has been shown in Sections 2.3 and 2.4 that the low-complexity implementation of the decision-feedback detector is possible based on the SQRD algorithm. For a variety of channel models, the SQRD detector is as good as the optimum V-BLAST detector and has the benefit of an order of magnitude smaller computing burden, being in this sense equivalent to the linear schemes. Thus, decision-feedback performance gain is achieved with relatively little implementation cost increase as compared to the linear detectors.

Any coherent detection scheme relies on explicit channel state information, which can be reliably estimated at the receiver by means of the pilot-assisted transmission methods. It has been discussed in Chapter 3 that the channel estimator design depends on the selected channel model, more specifically on its deterministic and statistical properties. We emphasise low-complexity extendable filters for the pilot-assisted channel estimation, which are implemented inside the transform-domain structure. Such an approach enables scalability of the pilot pattern and hence easy reconfigurability of pilot-aided transceivers depending on bandwidth and power availability.

The MMSE estimation algorithms can claim optimality only in the channels modelled as stationary stochastic processes and only in the presence of complete information about the channel response correlation and the noise variance. Incomplete information about the second-order statistics necessitates development of robust suboptimal estimation algorithms, which may not be as efficient as MMSE ones. It has been shown that the robust estimator design is justified in the case of interblock channel response filtering. The intrablock filtering schemes can greatly benefit from the knowledge of the channel impulse response (CIR) correlation by adopting MMSE smoothing; otherwise the simple intrablock CLS estimator is robust and sufficient.

The idea of adaptive learning of the channel statistics leads to recursive estimator implementations. In Subsections 3.3.2 and 3.4.5 we have introduced 1D and 2D recursive channel estimation algorithms. It has also been established by means of simulation examples (Subsection 3.6.3) that the recursive structures exhibit the most rapid convergence in the sample-spaced multipath channels. For this type of channels, the RLS-MMSE estimator architecture is sufficient to achieve maximum 2D filtering performance (i.e. the more advanced RMMSE-MMSE scheme does not have performance gain with respect to RLS-MMSE). It should also be noted that MIMO systems provide more possibilities for the channel statistics estimation than the SISO systems. This facilitates more efficient adaptive estimator implementations in the MIMO receivers.

The training (pilot) signals do not transfer any useful information. Their inclusion in the system bandwidth is necessary only to configure the receiver matching distorted modulated waveform at the radio channel output (channel estimation and equalisation stage). It is a trade-off between the accuracy of the channel distortion compensation and the binary data demodulation on the noise-affected constellation that governs bandwidth and power allocation between the pilot and data-modulated symbols. In this thesis, we have explicitly addressed the problem of pilot-data power allocation, assuming a general case of arbitrarily arranged pilot subcarriers, any spatial multiplexing order, and deployment-specific dynamic model of the channel. For several types of the practically feasible (either in the fixed or recursive mode) channel estimation algorithms, the suboptimal analytical expressions of the pilot-to-data power ratio (PDR) have been derived in Chapter 4. The practical importance of these mathematical results is the simplicity of the optimal PDR prediction for the given transceiver configuration and the known (or acquired) channel statistics information. In Section 4.3, it has been shown by a variety of numerical simulation examples that the proposed suboptimal PDR lies quite close to the actual optimal PDR value in most scenarios. But, in contrast to the introduced suboptimal PDR equations, the actual PDR cannot be determined analytically in the case of the complex system design, e.g., the MIMO-OFDM system operating in the doubly selective fading channel. The biggest analysis challenge in this case is the complexity of the channel estimation gain function.

5.2 Problems for future research

The findings presented in this thesis emphasise low-complexity techniques for the pilot-assisted channel estimation and spatially multiplexed data detection. Narrowing the scope of interest to the aforementioned areas has excluded or simplified a number of important transceiver design aspects in the thesis content. In this section, some of these aspects will be discussed in more detail, laying the foundation for future extensions to the current project. In particular, in the proposed research directions we continue to consider pilot-assisted multicarrier signal processing in the perfectly synchronised baseband system without any channel knowledge at the transmitter. We will begin with the straightforward extensions, which do not require alterations to the mathematical model of the system, used throughout this thesis. This will be followed by the more critical modifications (e.g., selection of a different pilot structure and adoption of more advanced channel estimation algorithms), necessitating analytical reformulation of the pilot-assisted transmission model.

5.2.1 Interblock interpolation of the channel response at the receiver

Strong channel response correlation between subsequent blocks makes it possible to include pilot symbols only in periodically transmitted blocks rather than in every block in a sequence. Higher correlation levels, resulting from small maximum Doppler shifts, allow for higher pilot periodicity factors (refer to Subsection 3.4.2 for details) and hence larger capacity gains. This property is extremely attractive for slowly fading fixed wireless access systems, for which the low pilot block rate releases effective bandwidth that creates an enormous potential for enlarging transmission capacity by means of spatial multiplexing (SM). In fact, the SM order is theoretically unlimited on the interval of the non-pilot-bearing blocks, for which the channel response is acquired based on pilot measurements obtained from the pilot-carrying blocks.

The channel response for non-pilot-bearing blocks can be estimated by interpolating the channel responses of adjacent pilot-bearing blocks. In the transform-domain receiver architecture, which simplifies channel estimation processing due to the reduced signal dimensions, the low-complexity interblock interpolator should be incorporated after the CIR filtering (CLS, CMMSE, CLS-MMSE, CMMSE-MMSE, etc.). The interpolator can be designed in matrix product form, by capitalising on the optimal rank reduction of the interblock CIR correlation matrix. There is however one major challenge in this approach: the optimal rank, defining the interpolation error variance, depends on the post-filtering noise variance (i.e. channel estimation MSE), which is difficult to calculate analytically. A robust solution, yielding a suboptimal rank setting and the corresponding pilot periodicity factor, could be a potential research target.

5.2.2 Non-sample-spaced channel estimation

If the multipath intensity profile of the channel includes components, which have non-sample-spaced excess delays with respect to the primary path, band-limiting at the receiver results in the CIR energy being spread over the block interval. In the system model adopted in the thesis, we have assumed that the CIR energy is compactly concentrated within the starting samples (refer to Subsection 2.2.1 for details), due to the fast decay of the sidelobes of the interpolating (band-limiting) function. Such an assumption allows the use of the Fourier basis to identify the lowpass CIR part, which can be decoupled from the highpass part known to contain only noise. For the non-sample-spaced channels such a transform cannot be achieved using the Fourier basis. Instead one needs to utilise the eigenbasis, which is specific to the considered channel correlation model. In most cases, however, the eigenbasis is unknown as the channel statistics information is not available at the receiver. Although adaptive learning of the channel statistics could be performed, this would lead to prohibitive latencies and substantially complicate the receiver architecture. In the absence of the actual channel eigenbasis, development of robust suboptimal transform bases, minimising CIR approximation error, is of interest (e.g., wavelets). Here it should be noted that not all bases are profitable from the standpoint of noise filtering at the post-transformation stage. Thus, some kind of optimisation (analytical or numerical) is necessary to determine transform basis, which would allow achieving the best estimation performance.

5.2.3 Pilot structure increasing MIMO transmission capacity

As pointed out in Section 4.1, SM-MIMO capacity of the pilot-carrying blocks is limited by the number of Tx antennas, each of which must have an independent pilot pattern. Furthermore, if every transmitted block incorporates pilot symbols, there exist an optimal number of Tx antennas, given by (4.2), that maximises transmission capacity.

It has been highlighted in Section 4.1 that for the SM-MIMO-OFDM systems operating in the slow fading propagation environments, which allow for the larger pilot periodicity coefficients (greater than the number of Tx antennas), a different training structure is recommended. It comprises periodical bursts of the successive pilot blocks, where each block in the burst carries a set of pilot subcarriers assigned to the individual Tx antenna. Thus, the capacity loss is observed only in the bandwidth portion occupied by a single antenna pilot subcarrier set (other antennas may not transmit at these frequencies). Higher pilot periodicity factors increase spectral efficiency, but lead to less accurate channel estimation because of the interblock CIR interpolation error. This trade-off motivates for the search of the optimal pilot periodicity as a function of the channel Doppler statistics and SNR. The analytical solution of the pilot periodicity problem for the selected channel estimation and interpolation methods represents an important contribution to both the theoretical capacity assessment and the practical design of the pilot-assisted SM-MIMO-OFDM system.

5.2.4 Optimal pilot structure and channel estimator design for cognitive radio

So far the mathematical model of the system, considered in this thesis, has been constrained by the white noise assumption and the preset pilot symbol positions in the time-frequency grid. These features are essential if the system is designed to operate in the dedicated frequency band. Potential application of multicarrier interfaces in the cognitive radio (CR) systems poses a number of critical alterations to the channel model, making the aforementioned assumptions no longer valid.

CR systems are designed to cope with the presence of interference from other band users. A CR transmitter utilises only parts of the frequency spectrum, which are not occupied by the primary users (e.g., the licenced wireless access terminals, operating in the given frequency band). However, there is always adjacent channel interference from the primary users to the secondary (cognitive) users, which originates from the spectrum side-lobes. The presence of this additive interference imposes coloured structure on the Gaussian noise model, which is now characterised by non-constant variance across the band and possibly non-zero covariance between the noise samples corresponding to different frequency bins. Thus, the channel estimation algorithms presented in Chapter 3 are no longer optimal for the CR scenarios. If the statistics of the background noise are known (or accurately measured by the spectrum-sensing mechanisms), channel estimators based on the LS and MMSE criteria can be modified to yield optimal performance.

A more serious problem in the CR application of the pilot-assisted multicarrier system is that the pilot structure cannot be fixed by design anymore. Thus, one has to perform adaptive loading of the pilot symbols in the time-frequency grid. Selection of a specific position for a pilot symbol would in general depend on the availability of this slot within the considered bandwidth at a given time instant (i.e. when the corresponding frequency bin on the given block interval is not occupied by a primary user transmission), quality (additive noise variance) and practical limitations (e.g., presence of the virtual subcarriers). One can see that a cumulative impact of all these factors leads to a novel complex optimisation problem in the wireless communication systems context – adaptive training loading. It is interesting to note that in contrast to the conventional adaptive data loading, used to increase transmission capacity by matching the instantaneous channel conditions, adaptive training loading requires no or negligible channel information signalling between the transmitter and the receiver. An analytical solution of the pilot loading problem can be searched for by trying to minimise channel estimation MSE. Based on the number of degrees of freedom in transmit signal adaptation, each individual pilot symbol can be assigned an appropriate position in the time-frequency grid and a transmit power so as to maximise channel estimator performance at the receiver side. The major disadvantage of the adaptive training pattern is the necessity to perform reconfiguration of the channel estimation scheme every time the pilot structure changes. Thus, CR flexibility comes at a price of substantially increased transceiver complexity.

References

[1] Cisco Systems, "Global IP traffic forecast and methodology, 2006-2011," *White Paper*, Jan. 2008.

[2] J. A. C. Bingham, "Multicarrier modulations for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5-14, May 1990.

[3] ETSI TS 300 401, *Radio broadcasting systems; Digital audio broadcasting (DAB) to mobile, portable and fixed receivers*, May 1997.

[4] M. Engels (Ed.), *Wireless OFDM Systems: How to Make Them Work?* Kluwer Academic Publishers, Boston, Dordrecht, London, 2002.

[5] ETSI EN 300 744, *Digital video broadcasting (DVB): Framing structure, channel coding and modulation for digital terrestrial television*, Nov. 2005.

[6] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, vol. 33, pp. 100-109, Feb. 1995.

[7] IEEE Std 802.11a, *Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer in the 5 GHz band*, 1999.

[8] R. van Nee, G. Awater, M. Morikura, H. Takanashi, M. Webster, and K. W. Halford, "New high-rate wireless LAN standards," *IEEE Commun. Mag.*, vol. 37, pp. 82-88, Dec. 1999.

[9] IEEE Std 802.11g, *Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: High-speed physical layer in the 5 GHz band; Amendment 4: Further higher data rate extension in the 2.4 GHz band*, 2003.

[10] ETSI TS 101 475, *Broadband Radio Access Networks (BRAN); HIPERLAN Type 2 technical specification: Part 1 – Physical (PHY) layer*, Oct. 1999.

[11] L. Yang, and G. B. Giannakis, "Ultra-wideband communications: An idea whose time has come," *IEEE Signal Process. Mag.*, vol. 21, pp. 26-54, Nov. 2004.

[12] J. S. Chow, J. C. Tu, and J. M. Cioffi, "A discrete multitone transceiver system for HDSL application," *IEEE J. Select. Areas Commun.*, vol. 9, pp. 895-908, Aug. 1991.

[13] K. Maxwell, "Asymmetric digital subscriber line: Interim technology for the next forty years," *IEEE Commun. Mag.*, vol. 34, pp. 100-106, Oct. 1996.

[14] G. Cherubini, E. Eleftheriou, S. Ölçer, and J. M. Cioffi, "Filter bank modulation techniques for very high-speed digital subscriber lines," *IEEE Commun. Mag.*, vol. 38, pp. 98-104, May 2000.

[15] L. Hanzo, C. H. Wong, and M. S. Yee, *Adaptive Wireless Transceivers*, Wiley, 2002.

[16] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.

[17] H. Sampath, S. Talwar, J. Tellado, V. Erceg, and A. Paulraj, "A fourth-generation MIMO-OFDM broadband wireless system: Design, performance and field trial results," *IEEE Commun. Mag.*, vol. 40, pp. 143-149, Sep. 2002.

[18] H. Yang, "A road to future broadband wireless access: MIMO-OFDM-based air interface," *IEEE Commun. Mag.*, vol. 43, pp. 53-60, Jan. 2005.

[19] H. Bolcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.

[20] IEEE Std 802.16-2004, *Part 16: Air interface for fixed broadband wireless access systems*, Oct. 2004.

[21] A. Ghosh, D. R. Wolter, J. G. Andrews, and R. Chen, "Broadband wireless access with WiMax/802.16: Current performance benchmarks and future potential," *IEEE Commun. Mag.*, vol. 43, pp. 129-136, Feb. 2005.

[22] IEEE Std 802.16-2005, *Part 16: Air interface for fixed and mobile broadband wireless access systems; Amendment 2: Physical and medium access control layers for combined fixed and mobile operation in licensed bands and Corrigendum 1*, Feb. 2006.

[23] F. Wang, A. Ghosh, C. Sankaran, P. J. Fleming, F. Hsieh, and S. J. Benes, "Mobile WiMAX systems: Performance and evolution," *IEEE Commun. Mag.*, vol. 46, pp. 41-49, Oct. 2008.

[24] A. Greenspan, M. Klerer, J. Tomcik, R. Canchi, and J. Wilson, "IEEE 802.20: Mobile broadband wireless access for the twenty-first century," *IEEE Commun. Mag.*, vol. 46, pp. 56-63, Jul. 2008.

[25] H. Ekstrom, A. Furuskar, J. Karlsson, M. Meyer, S. Parkvall, J. Torsner, and M. Wahlqvist, "Technical solutions for the 3G long-term evolution," *IEEE Commun. Mag.*, vol. 44, pp. 38-45, Mar. 2006.

[26] T. Paul, and T. Ogunfunmi, "Wireless LAN comes of age: Understanding the IEEE 802.11n amendment," *IEEE Circuits Syst. Mag.*, vol. 8, pp. 28-54, 1st Quarter 2008.

[27] S. Hara, and R. Prasad, "Overview of multicarrier CDMA," *IEEE Commun. Mag.*, vol. 35, pp. 126-133, Dec. 1997.

[28] Y. Zhou, T.-S. Ng, J. Wang, K. Higuchi, and M. Sawahashi, "OFCDM: A promising broadband wireless access technique," *IEEE Commun. Mag.*, vol. 46, pp. 38-49, Mar. 2008.

[29] Z. Wang, and G. B. Giannakis, "Wireless multicarrier communications: Where Fourier meets Shannon," *IEEE Signal Process. Mag.*, vol. 17, pp. 29-48, May 2000.

[30] T. Frank, A. Klein, and E. Costa, "IFDMA: A scheme combining the advantages of OFDMA and CDMA," *IEEE Wireless Commun.*, vol. 14, pp. 9-17, Jun. 2007.

[31] B. Farhang-Boroujeny, and R. Kempter, "Multicarrier communication techniques for spectrum sensing and communication in cognitive radios," *IEEE Commun. Mag.*, vol. 46, pp. 80-85, Apr. 2008.

[32] T. A. Weiss, and F. K. Jondral, "Spectrum Pooling: An Innovative Strategy for the Enhancement of Spectrum Efficiency," *IEEE Commun. Mag.*, vol. 42, pp. S8-14, Mar. 2004.

[33] G. Gu, J. He, X. Cao, and M. Naraghi-Pour, "An analytic approach to modeling and estimation of OFDM channels," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM),* pp. 2381-2386, Nov.-Dec 2004.

[34] J. K. Moon, and S. I. Choi, "Performance of channel estimation methods for OFDM systems in a multipath fading channel," *IEEE Trans. Consumer Electron.*, vol. 46, pp. 161-170, Feb. 2000.

[35] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, pp. 223-229, Sep. 2002.

[36] Y. Mostofi, and D. C. Cox, "Average error rate analysis for pilot-aided OFDM receivers with frequency-domain interpolation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, pp. 1421-1425, Mar. 2004.

[37] L. Deneire, P. Vandenameele, L. van der Perre, B. Gyselinckx, and M. Engels, "A low complexity ML channel estimator for OFDM communications," *IEEE Trans. Commun.*, vol. 51, pp. 135-140, Feb. 2003.

[38] M. Morelli, and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Process.*, vol. 49, pp. 3065-3073, Dec. 2001.

[39] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Upper Saddle River, NJ: Prentice Hall, 1993.

[40] J.-J. van be Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, pp. 815-819, Jul. 1995.

[41] H. Minn, and V. K. Bhargava, "An investigation into time-domain approach for OFDM channel estimation," *IEEE Trans. Broadcast.*, vol. 46, pp. 240-248, Dec. 2000.

[42] E. Golovins, and N. Ventura, "Modified order-recursive least squares estimator for the noisy OFDM channels," in *Proc. 5th IEEE Commun. Netw. Services Research (CNSR) conf.*, pp. 93-100, May 2007.

[43] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, pp. 931–939, Jul. 1998.

[44] H. Senol, H. A. Cirpan, and E. Panayirci, "Pilot-aided Bayesian MMSE channel estimation for OFDM systems: Algorithm and performance analysis," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM),* pp. 2361-2365, Nov.-Dec 2004.

[45] M. Noh, Y. Lee, and H. Park, "Low complexity LMMSE channel estimation for OFDM," *IEE Proc.-Commun.*, vol. 153, pp. 645-650, Oct. 2006.

[46] L. Huang, G. Mathew, and J. W. M. Bergmans, "Pilot-aided channel estimation for systems with virtual carriers," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, pp. 3070-3075, Jun. 2006.

[47] B. Yang, Z. Cao, and K. B. Letaief, "Analysis of low-complexity windowed DFT-based MMSE channel estimator for OFDM systems," *IEEE Trans. Commun.*, vol. 49, pp. 1977-1987, Nov. 2001.

[48] Y.-H. Yeh, and S.-G. Chen, "DCT-based channel estimation for OFDM systems," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, pp. 2442-2446, Jun. 2004.

[49] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay-subspace," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 315-325, Jan. 2004.

[50] H. Zamiri-Jafarian, H. Khoshbin, and S. Pasupathy, "Time-domain equalizer for OFDM systems based on SINR maximization," *IEEE Trans. Commun.*, vol. 53, pp. 924-930, Jun. 2005.

[51] P. J. W. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Trans. Commun.*, vol. 44, pp. 1662-1672, Dec. 1996.

[52] D. Daly, C. Heneghan, and A. D. Fagan, "Minimum mean-squared error impulse response shortening for discrete multitone transceivers," *IEEE Trans. Signal Process.*, vol. 52, pp. 301-306, Jan. 2004.

[53] N. Al-Dhahir, and J. M. Cioffi, "Optimum finite-length equalisation for multicarrier transceivers," *IEEE Trans. Commun.*, vol. 44, pp. 56-64, Jan. 1996.

[54] K. V. Acker, G. Leus, M. Moonen, O. van de Wiel, and T. Pollet, "Per-tone equalization for DMT-based systems," *IEEE Trans. Commun.*, vol. 49, pp. 109-119, Jan. 2001.

[55] D. Marelli, and M. Fu, "A subband approach to channel estimation and equalization for DMT and OFDM systems," *IEEE Trans. Commun.*, vol. 53, pp. 1850-1858, Nov. 2005.

[56] Y. (G.) Li, L. J. Cimini Jr., and N.R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 902-915, Jul. 1998.

[57] Y. (G.) Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 1207-1215, Jul. 2000.

[58] M. Sandell, and O. Edfors, *A Comparative Study of Pilot-Based Channel Estimators for Wireless OFDM*, Research Report TULEA 1996:19, Div. of Signal Processing, Lulea University of Technology, Sep. 1996.

[59] X. Wang, and K. J. R. Liu, "Channel estimation for multicarrier modulation systems using a time-frequency polynomial model," *IEEE Trans. Commun.*, vol. 50, pp. 1045-1049, Jul. 2002.

[60] M.-X. Chang, and Y. T. Su, "Model-based channel estimation for OFDM signals in Rayleigh fading," *IEEE Trans. Commun.*, vol. 50, pp. 540-545, Apr. 2002.

[61] X. Dong, W.-S. Lu, and A. C. K. Soong, "Linear interpolation in pilot symbol assisted channel estimation for OFDM," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1910-1920, May 2007.

[62] W. Chen, and R. Zhang, "Estimation of time and frequency selective channels in OFDM systems: A Kalman filter structure," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pp. 800-803, Nov.-Dec. 2004.

[63] W. H. He, and Y. Lee, "Low-complexity Kalman channel estimator structures for OFDM systems with and without virtual carriers," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, pp. 2447-2451, Jun. 2004. .

[64] J. Cai, X. Shen, and J. W. Mark, "Robust channel estimation for OFDM wireless communication systems – An H_∞ Approach," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 2060-2071, Nov. 2004.

[65] J. D. Parsons, *The Mobile Radio Propagation Channel*, 2nd ed., Wiley, 2000.

[66] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, pp. 467-479, Mar. 2001.

[67] M. Wax, and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, pp. 387-392, Apr. 1985.

[68] R. Roy, and T. Kailath, "ESPRIT – Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984-995, Jul. 1989.

[69] J. Liu, and X. Liu, "Time-varying channel identification and prediction in OFDM systems using 2-D frequency estimation," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, pp. 1-7, Oct. 2006.

[70] P. Chen, and H. Kobayashi, "Maximum likelihood channel estimation and signal detection for OFDM systems," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, pp. 1640-1645, Apr. 2002.

[71] J.-H. Park, M.-K. Oh, and D.-J. Park, "New channel estimation exploiting reliable decision-feedback symbols for OFDM systems," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, pp. 3046-3051, Jun. 2006.

[72] X. Wang, and K. J. R. Liu, "Performance analysis for adaptive channel estimation exploiting cyclic prefix in multicarrier modulation systems," *IEEE Trans. Commun.*, vol. 51, pp. 94-105, Jan. 2003.

[73] B. Han, X. Gao, X. You, and J. Wang, "An iterative joint channel estimation and symbol detection algorithm applied in OFDM system with high data to pilot power ratio," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, pp. 2076-2080, May 2003.

[74] S. Kalyani, and K. Giridhar, "Extreme value theory based decision directed OFDM channel tracking," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, pp. 2893-2898, Jun. 2006.

[75] S. Kalyani, and K. Giridhar, "Leverage weighted decision directed channel tracking for OFDM systems," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, pp. 2899-2904, Jun. 2006.

[76] C. K. Ho, B. Farhang-Boroujeny, and F. Chin, "Added pilot semi-blind channel estimation scheme for OFDM in fading channels," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pp. 3075-3079, Nov. 2001.

[77] S. Balasubramanian, B. Farhang-Boroujeny, and V. J. Mathews, "Pilot embedding for channel estimation and tracking in OFDM systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pp. 1244-1248, Nov.-Dec. 2004.

[78] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, pp. 1699-1712, Jul. 2002.

[79] C. Li, and S. Roy, "Subspace-based blind channel estimation for OFDM by exploiting virtual carriers," *IEEE Trans. Wireless Commun.*, vol. 2, pp 141-150, Jan. 2003.

[80] R. W. Heath, and G. B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Process.*, vol. 47, pp. 848-856, Mar. 1999.

[81] M. Luise, R. Reggiannini, and G. M. Vitetta, "Blind equalization/detection for OFDM signals over frequency-selective channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1568–1578, Oct. 1998.

[82] M.-X. Chang, and Y. T. Su, "Blind and semiblind detections of OFDM signals in fading channels," *IEEE Trans. Commun.*, vol. 52, pp. 744-754, May 2004.

[83] T. Cui, and C. Tellambura, "Joint data detection and channel estimation for OFDM systems," *IEEE Trans. Commun.*, vol. 54, pp. 670-679, Apr. 2006.

[84] S. Zhou, and G. B. Giannakis, "Finite-alphabet based channel estimation for OFDM and related multicarrier systems," *IEEE Trans. Commun.*, vol. 49, pp. 1402-1414, Aug. 2001.

[85] A. Petropulu, R. Zhang, and R. Lin, "Blind OFDM channel estimation through simple linear precoding," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 647-655, Mar. 2004.

[86] F. Gao, and A. Nallanathan, "Blind channel estimation for OFDM systems via a generalized precoding," *IEEE Trans. Veh. Technol.*, vol. 56, pp. 1155-1164, May 2007.

[87] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: General model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, pp. 12-25, Nov. 2004.

[88] R. Negi, and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consumer Electron.*, vol. 44, pp. 1122-1128, Aug. 1998.

[89] J. H. Manton, "Optimal training sequences and pilot tones for OFDM systems," *IEEE Commun. Lett.*, vol. 5, pp. 151-153, Apr. 2001.

[90] S. Ohno, and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with

application to wireless OFDM," *IEEE Trans. Commun.*, vol. 50, pp. 2113-2123, Dec. 2002.

[91] X. Cai, and G. B. Giannakis, "Error probability minimizing pilots for OFDM with M-PSK modulation over Rayleigh-fading channels," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 146-155, Jan. 2004.

[92] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, pp. 1351-1366, May 2003.

[93] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2338-2353, Aug. 2002.

[94] S. Ohno, and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 2138-2145, Sep. 2004.

[95] M. Dong, L. Tong, and B. M. Sadler, "Optimal pilot placement for channel tracking in OFDM," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, pp. 602-606, Oct. 2002.

[96] O. Simeone, and U. Spagnolini, "Adaptive pilot pattern for OFDM systems," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, pp. 978-982, Jun. 2004.

[97] J.-W. Choi, and Y.-H. Lee, "Optimum pilot pattern for channel estimation in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2083-2088, Sep. 2005.

[98] W. Zhang, X.-G. Xia, and P. C. Ching, "Optimal training and pilot pattern design for OFDM systems in Rayleigh fading," *IEEE Trans. Broadcast.*, vol. 52, pp. 505-514, Jun. 2006.

[99] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, pp. 1615-1623, Jun. 2003.

[100] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 453-465, Mar. 2005.

[101] B. Stantchev, and G. Fettweis, "Time-variant distortions in OFDM," *IEEE Commun. Lett.*, vol. 4, pp. 312-314, Sep. 2000.

[102] M. Brookes, "The Matrix Reference Manual," [online]: http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/into.html, 2005.

[103] A. D. Poularikas, *The Handbook of Formulas and Tables for Signal Processing*, CRC Press LLC, 1999.

[104] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology, and Techniques*, 2nd ed., Kluwer Academic/Plenum Publishers, 2000.

[105] D. M. J. Devasirvatham, "Multipath time delay spread in the digital portable radio environment," *IEEE Commun. Mag.,* vol. 25, pp. 13-21, Jun. 1987.

[106] ANSI J-STD-008, *Personal Station – Base Station Compatibility Requirements for 1.8 to 2.0 GHz Code Division Multiple Access (CDMA) Personal Communication Systems*, Mar. 1995.

[107] COST 207, *Digital Land Mobile Radio Communications*, Office for Official Publications of the European Communities, Final report, Luxembourg, 1989.

[108] A. A. M. Saleh, R. A. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 128–137, Feb. 1987.

[109] H. Hashemi, "Impulse response modeling of indoor radio propagation channels," *IEEE J. Select. Areas*

Commun., vol. 11, pp. 967–978, Sep. 1993.

[110] V. Erceg, et al., "A model for the multipath delay profile of fixed wireless channels," *IEEE J. Select. Areas Commun*., vol. 17, pp. 390–410, Mar. 1999.

[111] S. S. Ghassemzadeh, L. J. Greenstein, T. Sveinsson, and V. Tarokh, "A multipath intensity profile model for residential environments," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC),* pp. 150-155, Mar. 2003.

[112] L. Dossi, G. Tartara, and F. Tallone, "Statistical analysis of measured impulse response functions of 2.0 GHz indoor radio channels," *IEEE J. Select. Areas Commun*., vol. 14, pp. 405–410, Apr. 1996.

[113] R. J. C. Bultitude, P. Melançon, H. Zaghloul, G. Morrison, and M. Prokki, "The dependence of indoor radio channel multipath characteristics on transmit/receive ranges," *IEEE J. Select. Areas Commun*., vol. 11, pp. 979–990, Sep. 1993.

[114] S. J. Howard, and K. Pahlavan, "Autoregressive modeling of wide-band indoor radio propagation," *IEEE Trans. Commun.*, vol. 40, pp. 1540–1552, Sep. 1992.

[115] C. Cheon, G. Liang, and H. L. Bertoni, "Simulating radio channel statistics for different building environments," *IEEE J. Select. Areas Commun*., vol. 19, pp. 2191–2200, Nov. 2001.

[116] M. L. Rubio, A. Garcia-Armada, R. P. Torres, and J. L. Garcia, "Channel modeling and characterization at 17 GHz for indoor broadband WLAN," *IEEE J. Select. Areas Commun*., vol. 20, pp. 593–601, Nov. 2002.

[117] R. H. Clarke, and W. L. Khoo, "3-D mobile radio channel statistics," *IEEE Trans. Veh. Technol.*, vol. 46, pp. 798-799, Aug. 1997.

[118] W. C. Jakes, *Microwave Mobile Communications*, New York: Wiley, 1974.

[119] D. C. Cox, "Delay Doppler characteristics of multipath propagation at 910 MHz in a suburban mobile radio environment," *IEEE Trans. Antennas and Propagation*, vol. AP-20, pp. 625-635, Sep. 1972.

[120] S. Thoen, L. Van der Perre, and M. Engels, "Modeling the channel time-variance for fixed wireless communications," *IEEE Commun. Lett.*, vol. 6, pp. 331-333, Aug. 2002.

[121] A. Domazetovic, L. J. Greenstein, N. B. Mandayam, and I. Seskar, "Estimating the Doppler spectrum of a short-range fixed wireless channel", *IEEE Commun. Lett.*, vol. 7, pp. 227-229, May 2003.

[122] Y. (G.) Li, "Simplified channel estimation for OFDM systems with multiple transmit antennas," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 67-75, Jan. 2002.

[123] H. Yao, and G. W. Wornell, "Lattice-reduction-aided detectors for MIMO communication systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pp. 424-428, Nov. 2002.

[124] W. H. Mow, "Universal lattice decoding: a review and some recent results," in *Proc. IEEE Intern. Conf. Commun. (ICC)*, pp. 2842-2846, Jun. 2004.

[125] L. He, and H. Ge, "Reduced complexity maximum likelihood detection for V-BLAST systems," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, pp. 1386-1391, Oct. 2003.

[126] R. Bohnke, D. Wubben, V. Kuhn, and K.-D. Kammeyer, "Reduced complexity MMSE detection for BLAST architectures," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pp. 2258-2262, Dec. 2003.

[127] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1841-1852, Nov. 1999.

[128] S. W. Kim, "Log-likelihood ratio based detection ordering for the V-BLAST," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pp. 292-296, Dec. 2003.

[129] W. K. Wai, C.-Y. Tsui, and R. S. Cheng, "A low complexity architecture of the V-BLAST system," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2000.

[130] W. Yan, S. Sun, and Z. Lei, "A low complexity VBLAST OFDM detection algorithm for Wireless LAN systems," *IEEE Commun. Lett.*, vol. 8, pp. 374-376, Jun. 2004.

[131] K. M. Abadir, and J. R. Magnus, *Matrix Algebra: Econometric Exercises*, Cambridge University Press, 2005.

[132] M. N. Bandyopadhyay, *Introduction to Signals and Systems and DSP*, Prentice Hall of India, New Delhi, 2005.

[133] H. V. Sorensen, and C. S. Burrus, "Efficient computation of the DFT with only a subset of input or output points," *IEEE Trans. Signal Process.*, vol. 41, pp. 1184-1200, Mar. 1993.

[134] M. Z. A. Khan, S. A. Qadeer, and M. Y. Khan, "Fast computation of partial DFT for comb spectrum evaluation," *IEEE Signal Process. Lett.*, vol. 13, pp. 721-724, Dec. 2006.

[135] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Englewood Cliffs, NJ: Prentice-Hall, 1996.

[136] S. Song, J.-S. Lim, S. J. Baek, and K.-M. Sung, "Variable forgetting factor linear least squares algorithm for frequency selective fading channel estimation," *IEEE Trans. Veh. Technol.*, vol. 51, pp. 613-616, May 2002.

[137] D. Schafhuber, G. Matz, and F. Hlawatsch, "Adaptive Wiener filters for time-varying channel estimation in wireless OFDM systems," in *Proc. IEEE Intern. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, pp. 688-691, Apr. 2003.

[138] K. B. Petersen, and M. S. Pedersen, *The Matrix Cookbook*, Technical University of Denmark, Feb. 2006, [online] http://matrixcookbook.com.

[139] F. Gustafsson, *Adaptive Filtering and Change Detection*, Wiley, 2000.

[140] S. J. Elliott, and B. Rafaely, "Frequency-domain adaptation of causal digital filters," *IEEE Trans. Signal Process.*, vol. 48, pp. 1354-1364, May 2000.

[141] L. G. Ordonez, D. P. Palomar, A. Pages-Zamora, and J. R. Fonollosa, "High-SNR analytical performance of spatial multiplexing MIMO systems with CSI," *IEEE Trans. Signal Process.*, vol. 55, pp. 5447-5463, Nov. 2007.

[142] R. W. D. Nickalls, "A new approach to solving the cubic: Cardan's solution revealed," *The Mathematical Gazette*, vol. 77, pp. 354-359, 1993.

Appendix A

A.1: Dependence between CIR length and rms delay spread for sample-spaced exponential PDP

By substituting $\beta = B\tau_{\text{rms}}$, i.e. rms delay spread normalisation to the sampling rate, in (2.13) and using (2.14) and exponential series formulas [103],

$$\sum_{k=0}^{n} \mathrm{e}^{kb} = 1 + \mathrm{e}^{b}\frac{\mathrm{e}^{nb}-1}{\mathrm{e}^{b}-1} = \frac{1-\mathrm{e}^{(n+1)b}}{1-\mathrm{e}^{b}}, \tag{A.1}$$

$$\sum_{k=1}^{n} k\,x^{k} = \frac{x(1-x^{n})}{(1-x)^{2}} - \frac{n\,x^{n+1}}{1-x}, \tag{A.2}$$

$$\sum_{k=1}^{n} k^{2}x^{k} = \frac{x^{n+1}[-n^{2}x^{2} + (2n^{2}+2n-1)\,x - (n+1)^{2}] + x(x+1)}{(1-x)^{3}}, \tag{A.3}$$

where we let $b = -\alpha$, $x = \mathrm{e}^{-\alpha}$, $k = l$ and $n = L-1$, we can write down:

$$\begin{aligned}
\beta^{2} &= \frac{c_{\alpha}\sum_{l=0}^{L-1} l^{2}\,\mathrm{e}^{-\alpha l}}{c_{\alpha}\sum_{l=0}^{L-1}\mathrm{e}^{-\alpha l}} - \left(\frac{c_{\alpha}\sum_{l=0}^{L-1} l\,\mathrm{e}^{-\alpha l}}{c_{\alpha}\sum_{l=0}^{L-1}\mathrm{e}^{-\alpha l}}\right)^{2} \\
&= \frac{\left[\mathrm{e}^{-L\alpha}\left(-(L-1)^{2}\mathrm{e}^{-2\alpha} + (2(L-1)^{2} + 2(L-1) - 1)\mathrm{e}^{-\alpha} - L^{2}\right) + \mathrm{e}^{-\alpha}(1+\mathrm{e}^{-\alpha})\right]\left[1-\mathrm{e}^{-\alpha}\right]}{(1-\mathrm{e}^{-\alpha})^{3}(1-\mathrm{e}^{-L\alpha})} \\
&\quad - \left(\frac{\left[-(L-1)(1-\mathrm{e}^{-\alpha})\,\mathrm{e}^{-L\alpha} - \mathrm{e}^{-L\alpha} + \mathrm{e}^{-\alpha}\right]\left[1-\mathrm{e}^{-\alpha}\right]}{(1-\mathrm{e}^{-\alpha})^{2}(1-\mathrm{e}^{-L\alpha})}\right)^{2} \\
&= \frac{\mathrm{e}^{-L\alpha}\left[-(L-1)^{2}\mathrm{e}^{-2\alpha} + (2L^{2}-2L-1)\mathrm{e}^{-\alpha} - L^{2}\right]\left[1-\mathrm{e}^{-L\alpha}\right] + \mathrm{e}^{-\alpha}\left[1+\mathrm{e}^{-\alpha}\right]\left[1-\mathrm{e}^{-L\alpha}\right] - \left[((L-1)\mathrm{e}^{-\alpha} - L)\mathrm{e}^{-L\alpha} + \mathrm{e}^{-\alpha}\right]^{2}}{(1-\mathrm{e}^{-\alpha})^{2}(1-\mathrm{e}^{-L\alpha})^{2}}
\end{aligned}$$

$$= \frac{e^{-L\alpha}\left[(L-1)^2 e^{-(L+2)\alpha} - (2L^2-2L-1)e^{-(L+1)\alpha} + L^2 e^{-L\alpha} - (L-1)^2 e^{-2\alpha} + (2L^2-2L-1)e^{-\alpha} - L^2\right]}{(1-e^{-\alpha})^2(1-e^{-L\alpha})^2} +$$
$$+ \frac{e^{-\alpha}\left[-e^{-(L+1)\alpha} - e^{-L\alpha} + e^{-\alpha} + 1\right] - \left[((L-1)e^{-\alpha} - L)^2 e^{-2L\alpha} + 2((L-1)e^{-\alpha} - L)e^{-(L+1)\alpha} + e^{-2\alpha}\right]}{(1-e^{-\alpha})^2(1-e^{-L\alpha})^2}$$
$$= \frac{\left[(L-1)^2 - (L-1)^2\right]e^{-(2L+2)\alpha} + \left[-2L^2 + 2L + 1 + 2L(L-1)\right]e^{-(2L+1)\alpha} + (L^2 - L^2)e^{-2L\alpha}}{(1-e^{-\alpha})^2(1-e^{-L\alpha})^2} +$$
$$+ \frac{\left[-(L-1)^2 - 1 - 2(L-1)\right]e^{-(L+2)\alpha} + \left[2L^2 - 2L - 1 - 1 + 2L\right]e^{-(L+1)\alpha} - L^2 e^{-L\alpha} + (1-1)e^{-2\alpha} + e^{-\alpha}}{(1-e^{-\alpha})^2(1-e^{-L\alpha})^2}$$

$$= \frac{e^{-(2L+1)\alpha} - L^2 e^{-(L+2)\alpha} + 2(L^2-1)e^{-(L+1)\alpha} - L^2 e^{-L\alpha} + e^{-\alpha}}{(1-e^{-\alpha})^2(1-e^{-L\alpha})^2}$$
$$= \frac{e^{-(L+1)\alpha}\left[e^{L\alpha} + e^{-L\alpha} - L^2(e + e^{-\alpha}) + 2(L^2-1)\right]}{\left[e^{-(L+1)\alpha} - e^{-L\alpha} - e^{-\alpha} + 1\right]^2} \quad . \tag{A.4}$$
$$= \frac{e^{-(L+1)\alpha}\left[e^{L\alpha} + e^{-L\alpha} - L^2(e + e^{-\alpha}) + 2(L^2-1)\right]}{\left[e^{-(L+1)\alpha/2}(e^{-(L+1)\alpha/2} - e^{-(L-1)\alpha/2} - e^{(L-1)\alpha/2} + e^{(L+1)\alpha/2})\right]^2}$$

Using the hyperbolic trigonometry identities,

$$2\cosh u = e^{u} - e^{-u}, \tag{A.5}$$

$$\cosh u - \cosh v = 2\sinh\frac{(u+v)}{2}\sinh\frac{(u-v)}{2}, \tag{A.6}$$

$$2\sinh^2\frac{u}{2} = \cosh u - 1, \tag{A.7}$$

(A.4) reduces to

$$\beta^2 = \frac{2(\cosh L\alpha - L^2\cosh\alpha + L^2 - 1)}{4\left[\cosh\frac{(L+1)\alpha}{2} - \cosh\frac{(L-1)\alpha}{2}\right]^2}$$
$$= \frac{\cosh L\alpha - L^2\cosh\alpha + L^2 - 1}{8\sinh^2\frac{L\alpha}{2}\sinh^2\frac{\alpha}{2}} =$$
$$= \frac{\cosh L\alpha - L^2\cosh\alpha + L^2 - 1}{2(\cosh L\alpha - 1)(\cosh\alpha - 1)}$$
$$= \frac{1}{2(\cosh\alpha - 1)} - \frac{L^2}{2(\cosh L\alpha - 1)} \quad . \tag{A.8}$$

Hence, the mathematical relationship between the exponential factor α, CIR length L and the sample-normalised rms delay spread β for the finite-length exponential PDP has been obtained in the form of

$$\beta = \sqrt{\frac{1}{2(\cosh\alpha - 1)} - \frac{L^2}{2(\cosh L\alpha - 1)}}. \tag{A.9}$$

It is of interest to analyse (A.8) for the two marginal argument values, namely when $L \to \infty$ and $\alpha = 0$. In the following derivations it is shown that for these cases an inverse function exists in the closed form.

$$\begin{aligned}\lim_{L\to\infty}\beta^2 &= \lim_{L\to\infty}\frac{1}{2(\cosh\alpha-1)}-\frac{L^2}{2(\cosh L\alpha-1)}\\ &= \frac{1}{2(\cosh\alpha-1)}-\lim_{L\to\infty}\frac{\frac{\partial^2}{\partial L^2}L^2}{2\frac{\partial^2}{\partial L^2}(\cosh L\alpha-1)}\\ &= \frac{1}{4\sinh^2\frac{\alpha}{2}}-\lim_{L\to\infty}\frac{1}{\alpha^2\cosh L\alpha}\\ &= \left(2\sinh\frac{\alpha}{2}\right)^{-2}\end{aligned} \quad . \tag{A.10}$$

Hence

$$\beta_{L\to\infty} = \frac{1}{2\sinh\alpha/2}, \tag{A.11}$$

and consequently

$$\alpha_{L\to\infty} = 2\,\text{arcsin h}\frac{1}{2\beta_{L\to\infty}} = 2\ln\left(\frac{1}{2\beta_{L\to\infty}}+\sqrt{\frac{1}{4\beta_{L\to\infty}^2}+1}\right). \tag{A.12}$$

The simplest way to determine CIR length L, corresponding to $\alpha=0$ for the given β, is to make use of the initial rms delay spread definition (2.13), as well as the series sum equations [103]

$$\sum_{k=1}^{n}k = \frac{n(n+1)}{2}, \tag{A.13}$$

$$\sum_{k=1}^{n}k^2 = \frac{n(n+1)(2n+1)}{6}, \tag{A.14}$$

where $k=l$ and $n=L-1$. Hence

$$\begin{aligned}\beta_{\alpha=0}^2 &= L^{-1}\sum_{l=0}^{L-1}l^2-L^{-2}\left(\sum_{l=0}^{L-1}l\right)^2\\ &= \frac{L(L-1)(2L-1)}{6L}-\frac{L^2(L-1)^2}{4L^2}\\ &= \frac{2(L-1)(2L-1)-3(L-1)^2}{12}\\ &= \frac{(L-1)(L+1)}{12} = \frac{1}{12}\left(L^2-1\right)\end{aligned} \quad , \tag{A.15}$$

from where

$$\beta_{\alpha\to0} = \frac{1}{2}\sqrt{\frac{L^2-1}{3}}, \tag{A.16}$$

and the corresponding CIR length is calculated as

$$L_{\alpha\to0} = \sqrt{12\beta_{\alpha\to0}^2+1}\,. \tag{A.17}$$

A.2: Frequency correlation function for finite-length exponentially-decaying PDP

Taking DTFT of (2.14) one can obtain FCF in the form of

$$R(u)=\sum_{l=0}^{L-1}\frac{(1-\mathrm{e}^{-\alpha})}{(1-\mathrm{e}^{-L\alpha})}\mathrm{e}^{-\alpha l}\mathrm{e}^{-\mathrm{j}ul}=\frac{(1-\mathrm{e}^{-\alpha})}{(1-\mathrm{e}^{-L\alpha})}\sum_{l=0}^{L-1}\mathrm{e}^{-(\alpha+\mathrm{j}u)l}\ , \tag{A.18}$$

where $u\in[-\pi,\pi)$ denotes angular frequency normalised to the sampling rate. Using the sum of exponents equality from [103],

$$\sum_{k=0}^{n}\mathrm{e}^{k\,x}=\frac{1-\mathrm{e}^{(n+1)x}}{1-\mathrm{e}^{x}}=\frac{\sinh\frac{(n+1)\,x}{2}}{\sinh\frac{x}{2}}\mathrm{e}^{\frac{nx}{2}}\ , \tag{A.19}$$

where $x=-(\alpha+\mathrm{j}u)$, $k=l$ and $n=L-1$, (A.18) is reduced to

$$R(u)=\frac{(1-\mathrm{e}^{-\alpha})}{(1-\mathrm{e}^{-L\alpha})}\frac{(1-\mathrm{e}^{-L(\alpha+\mathrm{j}u)})}{(1-\mathrm{e}^{-(\alpha+\mathrm{j}u)})}=\frac{(1-\mathrm{e}^{-\alpha})}{(1-\mathrm{e}^{-L\alpha})}\mathrm{e}^{-\frac{(L-1)(\alpha+\mathrm{j}u)}{2}}\frac{\sinh\frac{L(\alpha+\mathrm{j}u)}{2}}{\sinh\frac{(\alpha+\mathrm{j}u)}{2}}. \tag{A.20}$$

In case of the uniform PDP ($\alpha=0$),

$$R_{\mathrm{u}}(u)=\lim_{\alpha\to 0}R(u)=\frac{1}{L}\frac{\sin\frac{Lu}{2}}{\sin\frac{u}{2}}\mathrm{e}^{-\frac{\mathrm{j}u(L-1)}{2}}. \tag{A.21}$$

A.3: Rms spreads for various Doppler spectra

The rms angular spread for the symmetric Doppler spectrum is expressed from (2.29) as

$$\omega_{\mathrm{rms}}=\frac{1}{2\pi}\sqrt{\frac{\int_0^{\pi}\omega^2 S(\omega)\,\mathrm{d}\omega}{\int_0^{\pi}S(\omega)\,\mathrm{d}\omega}}\ , \tag{A.22}$$

where the total power of $S(\omega)$ is normalised to unity, i.e. $\frac{1}{\pi}\int_0^{\pi}S(\omega)\,\mathrm{d}\omega=1$.

For the Flat PSD (2.30) it can be shown that

$$\omega_{\mathrm{rms,F}} = \frac{1}{2\pi}\sqrt{\frac{\int_0^{\omega_\mathrm{D}} \omega^2 \mathrm{d}\omega}{\int_0^{\omega_\mathrm{D}} \mathrm{d}\omega}} = \frac{1}{2\pi}\sqrt{\frac{\omega_\mathrm{D}^3}{3\omega_\mathrm{D}}} = \frac{\omega_\mathrm{D}}{2\sqrt{3}\pi}\,. \tag{A.23}$$

For the Jakes spectrum (2.33):

$$\omega_{\mathrm{rms,J}} = \frac{1}{2\pi}\sqrt{\frac{\int_0^{\omega_\mathrm{D}} \omega^2\left[\omega_\mathrm{D}^2-\omega^2\right]^{-1/2}\mathrm{d}\omega}{\int_0^{\omega_\mathrm{D}} \left[\omega_\mathrm{D}^2-\omega^2\right]^{-1/2}\mathrm{d}\omega}} = \frac{1}{2\pi}\sqrt{\frac{\frac{\omega_\mathrm{D}^2}{2}\arcsin\left(\frac{\omega_\mathrm{D}}{\omega_\mathrm{D}}\right)}{\pi/2}} = \frac{\omega_\mathrm{D}}{2\sqrt{2}\pi}\,, \tag{A.24}$$

where the integration identity $\int \omega^2\left[\omega_\mathrm{D}^2-\omega^2\right]^{-1/2}\mathrm{d}\omega = -\frac{\omega}{2}\sqrt{\omega_\mathrm{D}^2-\omega^2} + \frac{\omega_\mathrm{D}^2}{2}\arcsin\left(\frac{\omega}{\omega_\mathrm{D}}\right)$ [103] has been used.

For the Gaussian PSD (2.36):

$$\omega_{\mathrm{rms,G}} = \frac{1}{2\pi}\sqrt{\frac{1}{\pi}\frac{2\sqrt{\pi}}{\omega_\mathrm{D}}\int_0^{\pi} \omega^2 \exp\left[-\left(\frac{\omega}{\omega_\mathrm{D}}\right)^2\right]\mathrm{d}\omega} = \frac{\omega_\mathrm{D}}{2\pi}\sqrt{\frac{2}{\sqrt{\pi}}\int_0^{\pi/\omega_\mathrm{D}} u^2\exp\left[-u^2\right]\mathrm{d}u}\,. \tag{A.25}$$

As it is typically true that $\omega_\mathrm{D} << \pi$ (especially for the slow-varying channels), one can use the identity $\lim_{\pi/\omega_\mathrm{D}\to\infty}\int_0^{\pi/\omega_\mathrm{D}} u^2\exp\left[-u^2\right]\mathrm{d}u \to \frac{\sqrt{\pi}}{4}$ [103] to reduce (A.25) to $\omega_{\mathrm{rms,G}} \approx \frac{\omega_\mathrm{D}}{2\sqrt{2}\pi}$.

Appendix B

B.1: Example of matrix inverse computation using Gauss-Jordan elimination

Gauss-Jordan elimination is known as the most stable and efficient procedure to compute the matrix inverse [131]. To illustrate its application, consider an example of the non-singular matrix

$$\mathbf{A} = \begin{bmatrix} 3 & -2 & 1 \\ 1 & 2 & 2 \\ 4 & 5 & 6 \end{bmatrix}.$$

The key idea of the algorithm summarised in Tab.2.2 is to perform elementary row operations on the augmented matrix $[\mathbf{A}\,\mathbf{I}]$, where $\mathbf{I}$ is the square identity matrix, until one obtains $[\mathbf{I}\,\mathbf{A}^{-1}]$. In the following table these actions will be conducted step-by-step. We will also make remarks about the computational effort required.

No.	Matrix record	Computational load
1)	$\left[\begin{array}{ccc\|ccc} 3 & -2 & 1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 & 1 & 0 \\ 4 & 5 & 6 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} \\ row_2 - 1/3\, row_1 \\ row_3 - 4/3\, row_1 \end{array}$	3 multiplications and 2 additions on each of 2 rows (division operation is equivalent to multiplication operation)
2)	$\left[\begin{array}{ccc\|ccc} 3 & -2 & 1 & 1 & 0 & 0 \\ 0 & 8/3 & 5/3 & -1/3 & 1 & 0 \\ 0 & 23/3 & 14/3 & -4/3 & 0 & 1 \end{array}\right] \begin{array}{l} row_1 + 3/4\, row_2 \\ \\ row_3 - 23/8\, row_2 \end{array}$	3 multiplications and 2 additions on each of 2 rows
3)	$\left[\begin{array}{ccc\|ccc} 3 & 0 & 9/4 & 3/4 & 3/4 & 0 \\ 0 & 8/3 & 5/3 & -1/3 & 1 & 0 \\ 0 & 0 & -1/8 & -3/8 & -23/8 & 1 \end{array}\right] \begin{array}{l} row_1 + 18\, row_3 \\ row_2 + 40/3\, row_3 \\ \\ \end{array}$	3 multiplications and 2 additions on each of 2 rows

4)	$\left[\begin{array}{ccc\|ccc} 3 & 0 & 0 & -6 & -51 & 18 \\ 0 & 8/3 & 0 & -16/3 & -112/3 & 40/3 \\ 0 & 0 & -1/8 & -3/8 & -23/8 & 1 \end{array}\right] \begin{array}{l} 1/3\,row_1 \\ 3/8\,row_2 \\ -8\,row_3 \end{array}$	3 multiplications for each row
Result	$\left[\begin{array}{ccc\|ccc} 1 & 0 & 0 & -2 & -17 & 6 \\ 0 & 1 & 0 & -2 & -14 & 5 \\ 0 & 0 & 1 & 3 & 23 & -8 \end{array}\right]$	

Hence

$$\mathbf{A}^{-1} = \begin{bmatrix} -2 & -17 & 6 \\ -2 & -14 & 5 \\ 3 & 23 & -8 \end{bmatrix}.$$

Note that the number of mathematical operations at steps 1-3 remains constant. The only difference between the described example and the algorithm in Tab.2.2 is that in Tab.2.2 the elementary operations are performed along matrix columns.

Now consider a triangular matrix example:

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 4 & 5 \\ 0 & -2 & 3 & -4 \\ 0 & 0 & -1 & -3 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

No.	Matrix record	Computational load
1)	$\left[\begin{array}{cccc\|cccc} 1 & 3 & 4 & 5 & 1 & 0 & 0 & 0 \\ 0 & -2 & 3 & -4 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & -3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} \\ row_2 + 0\,row_1 \\ row_3 + 0\,row_1 \\ row_4 + 0\,row_1 \end{array}$	no multiplications and no additions needed for the first column
2)	$\left[\begin{array}{cccc\|cccc} 1 & 3 & 4 & 5 & 1 & 0 & 0 & 0 \\ 0 & -2 & 3 & -4 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & -3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 1 \end{array}\right] row_1 + 3/2\,row_2$	3 multiplications and 2 additions on 1 row
3)	$\left[\begin{array}{cccc\|cccc} 1 & 0 & 17/2 & -1 & 1 & 3/2 & 0 & 0 \\ 0 & -2 & 3 & -4 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & -3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} row_1 + 17/2\,row_3 \\ row_2 + 3\,row_3 \\ \\ \end{array}$	2 multiplications and 1 addition on each of 2 rows

4)	$\left[\begin{array}{cccc\|cccc} 1 & 0 & 0 & -53/2 & 1 & 3/2 & 17/2 & 0 \\ 0 & -2 & 0 & -13 & 0 & 1 & 3 & 0 \\ 0 & 0 & -1 & -3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} row_1 + 53/8\, row_4 \\ row_2 + 13/4\, row_4 \\ row_3 + 3/4\, row_4 \\ \end{array}$	1 multiplication and no additions on each of 3 rows
5)	$\left[\begin{array}{cccc\|cccc} 1 & 0 & 0 & 0 & 1 & 3/2 & 17/2 & 53/8 \\ 0 & -2 & 0 & 0 & 0 & 1 & 3 & 13/4 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 3/4 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 1 \end{array}\right] \begin{array}{l} 1\, row_1 \\ -1/2\, row_2 \\ -1\, row_3 \\ 1/4\, row_4 \end{array}$	4 multiplications for the first row, 3 – for the second, 2 – for the third and 1 – for the fourth
Result	$\left[\begin{array}{cccc\|cccc} 1 & 0 & 0 & 0 & 1 & 3/2 & 17/2 & 53/8 \\ 0 & 1 & 0 & 0 & 0 & -1/2 & -3/2 & -13/8 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & -3/4 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1/4 \end{array}\right]$	

Hence

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & 3/2 & 17/2 & 53/8 \\ 0 & -1/2 & -3/2 & -13/8 \\ 0 & 0 & -1 & -3/4 \\ 0 & 0 & 0 & 1/4 \end{bmatrix}.$$

For a general case of $n \times n$ matrix $\mathbf{A}$ the computational complexity is expressed as follows.

Total number of multiplications:

$$C_{\mathrm{M}} = n^2 + \sum_n n(n-1) = n^3. \tag{B.1}$$

Total number of additions:

$$C_{\mathrm{A}} = \sum_n (n-1)^2 = n^3 - 2n^2 + n. \tag{B.2}$$

If $\mathbf{A}$ is of triangular shape, then the complexity is about six times lesser.

Total number of multiplications:

$$C_{\mathrm{M}} = \sum_{i=1}^{n} i + \sum_{i=1}^{n-1} i\,(n-i) = n - \sum_{i=1}^{n-1} i^2 + (n+1)\sum_{i=1}^{n-1} i = n - \frac{n}{6}(n-1)(2n-1) + \frac{n}{2}(n-1)(n+1)$$
$$= n - \frac{2n^3 - 3n^2 + n}{6} + \frac{n^3 - n}{2} = \frac{n^3}{6} + \frac{n^2}{2} + \frac{n}{3} \tag{B.3}$$

Total number of additions:

$$C_{\mathrm{A}} = \sum_{i=1}^{n-2} i(n-i-1) = -\sum_{i=1}^{n-2} i^2 + (n-1)\sum_{i=1}^{n-2} i = -\frac{1}{6}(n-2)(n-1)(2n-3) + \frac{1}{2}(n-2)(n-1)^2$$
$$= -\frac{1}{6}(n^2-3n+2)(2n-3) + \frac{1}{2}(n^2-2n+1)(n-2)$$
$$= -\frac{1}{6}(2n^3-9n^2+13n-6) + \frac{1}{2}(n^3-4n^2+5n-2) = \frac{n^3}{6} - \frac{n^2}{2} + \frac{n}{3} \quad . \qquad \text{(B.4)}$$

B.2: Example of V-BLAST detection

Herein we assess computational complexity per subcarrier of the V-BLAST algorithm in its classic version (Tab.2.1), assuming ZF weighting. As one of $\tilde{\mathbf{H}}$ columns is zeroed per recursion, complexity of pseudoinverse computation in Tab.2.2 decreases with i increment, being proportional to $N_{\mathrm{tx}} - i$. In fact, at *i*th recursion the inversion is applied to a submatrix inside $\mathbf{K}$, which is of the size $(N_{\mathrm{tx}} - i) \times (N_{\mathrm{tx}} - i)$ and is formed by excluding zero elements in all the rows and columns indexed k_j, $j \in [0, i-1]$. After the inverse of this submatrix is computed, it is stuffed up to the initial size $N_{\mathrm{tx}} \times N_{\mathrm{tx}}$ with zeros in the same positions.

Consider an example of $N_{\mathrm{tx}} = 3$, $N_{\mathrm{rx}} = 4$ and $\mathbf{x}_{n,m} = [1\mathrm{i} \quad 1\mathrm{i} \quad -1]^{\mathrm{T}}$ being a vector of QPSK-modulated symbols loaded on the *n*th subcarrier inside the *m*th block. Let the MIMO transform matrix and the received symbols vector be respectively:

$$\mathbf{H}_{n,m} = \begin{bmatrix} 0.89 - 1.13\mathrm{i} & -1.02 + 0.4\mathrm{i} & -0.28 + 0.49\mathrm{i} \\ 0.58 + 0.5\mathrm{i} & 0.91 + 0.47\mathrm{i} & 0.84 - 0.85\mathrm{i} \\ -0.01 - 0.11\mathrm{i} & -1.13 + 0.18\mathrm{i} & -0.75 + \mathrm{i} \\ -0.57 + 0.37\mathrm{i} & 0.16 - 0.65\mathrm{i} & -1.53 - 0.04\mathrm{i} \end{bmatrix} \text{ and } \mathbf{y}_{n,m} = \begin{bmatrix} 1.12 - 0.76\mathrm{i} \\ -1.77 + 1.28\mathrm{i} \\ 0.51 - 1.99\mathrm{i} \\ 1.78 - 0.48\mathrm{i} \end{bmatrix}.$$

For comparison purposes the linear ZF solution (2.48), causing a detection error, is given first:

$$\hat{\mathbf{x}}_{n,m} = \mathbf{H}^{+}_{n,m}\mathbf{y}_{n,m} = \begin{bmatrix} -0.24 + 0.54\mathrm{i} \\ -0.69 + 0.62\mathrm{i} \\ -0.85 + 0.44\mathrm{i} \end{bmatrix}.$$

Demodulating ZF symbol estimates,

$$\tilde{\mathbf{x}}_{n,m} = \mathrm{demod}[\hat{\mathbf{x}}_{n,m}] = \begin{bmatrix} 1\mathrm{i} \\ -1 \\ -1 \end{bmatrix},$$

produces an error in the 2nd symbol.

Using equations (B.1) and (B.2), describing complexity of the pseudoinverse computation, we can express the total number of complex multiplications required by the conventional ZF detector as

$$C_{\mathrm{M}} = N_{\mathrm{tx}}^3 + N_{\mathrm{rx}}N_{\mathrm{tx}}^2 + N_{\mathrm{rx}}N_{\mathrm{tx}} = (N_{\mathrm{rx}} + N_{\mathrm{tx}})N_{\mathrm{tx}}^2 + N_{\mathrm{rx}}N_{\mathrm{tx}}, \qquad \text{(B.5)}$$

and the total number of additions as

$$C_{\mathrm{A}} = N_{\mathrm{tx}}^3 - 2N_{\mathrm{tx}}^2 + N_{\mathrm{tx}} + N_{\mathrm{rx}}N_{\mathrm{tx}}(N_{\mathrm{tx}} - 1) + (N_{\mathrm{rx}} - 1)N_{\mathrm{tx}} = (N_{\mathrm{rx}} + N_{\mathrm{tx}} - 2)N_{\mathrm{tx}}^2 . \quad \text{(B.6)}$$

Note that the MMSE detector with the weighting matrix (2.49) has the same multiplicative complexity as ZF and only N_{tx} operations higher additive complexity. Hence both ZF and MMSE linear detectors are characterised by the computational load of the order $O[(N_{\mathrm{rx}} + N_{\mathrm{tx}})N_{\mathrm{tx}}^2]$.

Now we apply ZF-VBLAST algorithm (Tab.2.1) to the same scenario. The algorithm starts with the initial variable assignments:

$$\tilde{\mathbf{y}} = \begin{bmatrix} 1.12 - 0.76\mathrm{i} \\ -1.77 + 1.28\mathrm{i} \\ 0.51 - 1.99\mathrm{i} \\ 1.78 - 0.48\mathrm{i} \end{bmatrix},\ \mathbf{A} = \mathbf{H}_{n,m}^{\mathrm{H}}\mathbf{H}_{n,m},\ \tilde{\mathbf{H}} = \mathbf{H}_{n,m},$$

and its subsequent steps are described as follows.

		Operation record	Computational load
Init.		$\tilde{\mathbf{y}} = \begin{bmatrix} 1.12 - 0.76\mathrm{i} \\ -1.77 + 1.28\mathrm{i} \\ 0.51 - 1.99\mathrm{i} \\ 1.78 - 0.48\mathrm{i} \end{bmatrix}$	
		$\mathbf{A} = \mathbf{H}_{n,m}^{\mathrm{H}}\mathbf{H}_{n,m} = \begin{bmatrix} 3.13 & -0.94 - 0.79\mathrm{i} & 0.01 - 0.3\mathrm{i} \\ -0.94 + 0.79\mathrm{i} & 4.01 & 1.66 - 3.55\mathrm{i} \\ 0.01 + 0.3\mathrm{i} & 1.66 + 3.55\mathrm{i} & 5.65 \end{bmatrix}$	3x4 multiplications for the 1st column, 2x4 - for the 2nd column and 1x4 - for the 3rd column; 3x3 additions for the 1st column, 2x3 - for the 2nd column and 1x3 - for the 3rd column
$i = 0$	Init.	$\tilde{\mathbf{H}} := \begin{bmatrix} 0.89 - 1.13\mathrm{i} & -1.02 + 0.4\mathrm{i} & -0.28 + 0.49\mathrm{i} \\ 0.58 + 0.5\mathrm{i} & 0.91 + 0.47\mathrm{i} & 0.84 - 0.85\mathrm{i} \\ -0.01 - 0.11\mathrm{i} & -1.13 + 0.18\mathrm{i} & -0.75 + \mathrm{i} \\ -0.57 + 0.37\mathrm{i} & 0.16 - 0.65\mathrm{i} & -1.53 - 0.04\mathrm{i} \end{bmatrix}$	
	(a)	$\mathbf{K} := \mathbf{A},\ \mathbf{L} := \mathbf{I}$	

		$r=0$	$\mathbf{k} := [\mathbf{K}]_{0,0}^{-1}(\mathbf{K}^{\mathrm{T}})_0 - (\mathbf{I})_0 = \begin{bmatrix} 0 \\ -0.3-0.25\mathrm{i} \\ 0-0.09\mathrm{i} \end{bmatrix}$	2 multiplications and 0 additions
			$\mathbf{K} := \mathbf{K} - (\mathbf{K})_0\mathbf{k}^{\mathrm{T}} = \begin{bmatrix} 3.13 & 0 & 0 \\ -0.94+0.79\mathrm{i} & 3.52 & 1.58-3.64\mathrm{i} \\ 0.01+0.3\mathrm{i} & 1.58+3.64\mathrm{i} & 5.62 \end{bmatrix}$	2 multiplications for the 2nd column, 2 - for the 3rd column; 2 additions for the 2nd column, 2 - for the 3rd column
			$\mathbf{L} := \mathbf{L} - (\mathbf{L})_0\mathbf{k}^{\mathrm{T}} = \begin{bmatrix} 1 & 0.3+0.25\mathrm{i} & 0.09\mathrm{i} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	0 multiplications and 0 additions
		$r=1$	$\mathbf{k} := [\mathbf{K}]_{1,1}^{-1}(\mathbf{K}^{\mathrm{T}})_1 - (\mathbf{I})_1 = \begin{bmatrix} -0.27+0.23\mathrm{i} \\ 0 \\ 0.45-1.03\mathrm{i} \end{bmatrix}$	2 multiplications and 0 additions
			$\mathbf{K} := \mathbf{K} - (\mathbf{K})_1\mathbf{k}^{\mathrm{T}} = \begin{bmatrix} 3.13 & 0 & 0 \\ 0 & 3.52 & 0 \\ 1.26+0.91\mathrm{i} & 1.58+3.64\mathrm{i} & 1.14 \end{bmatrix}$	1 multiplication for the 1st column, 1 - for the 3rd column; 1 addition for the 1st column, 1 - for the 3rd column
			$\mathbf{L} := \mathbf{L} - (\mathbf{L})_1\mathbf{k}^{\mathrm{T}} = \begin{bmatrix} 1.14 & 0.3+0.25\mathrm{i} & -0.4+0.29\mathrm{i} \\ 0.27-0.23\mathrm{i} & 1 & -0.45+1.03\mathrm{i} \\ 0 & 0 & 1 \end{bmatrix}$	1 multiplication for the 1st column, 1 - for the 3rd column; 1 addition for the 1st column, 1 - for the 3rd column
		$r=2$	$\mathbf{k} := [\mathbf{K}]_{2,2}^{-1}(\mathbf{K}^{\mathrm{T}})_2 - (\mathbf{I})_2 = \begin{bmatrix} 1.1+0.8\mathrm{i} \\ 1.39+3.19\mathrm{i} \\ 0 \end{bmatrix}$	2 multiplications and 0 additions
			$\mathbf{K} := \mathbf{K} - (\mathbf{K})_2\mathbf{k}^{\mathrm{T}} = \begin{bmatrix} 3.13 & 0 & 0 \\ 0 & 3.52 & 0 \\ 0 & 0 & 1.14 \end{bmatrix}$	0 multiplications and 0 additions

			$\mathbf{L} := \mathbf{L} - (\mathbf{L})_2 \mathbf{k}^{\mathrm{T}}$ $= \begin{bmatrix} 1.81 & 1.78+1.13\mathrm{i} & -0.4+0.29\mathrm{i} \\ 1.58-1\mathrm{i} & 4.92 & -0.45+1.03\mathrm{i} \\ -1.1-0.8\mathrm{i} & -1.39-3.19\mathrm{i} & 1 \end{bmatrix}$	2 multiplications for the 1st column, 2 - for the 2nd column; 2 additions for the 1st column, 2 - for the 2nd column
		$\mathbf{G} := (\mathbf{L}\mathbf{K}^{-1}\tilde{\mathbf{H}}^{\mathrm{H}})^{\mathrm{T}}$ $= \begin{bmatrix} 0.35+0.22\mathrm{i} & 0.44-0.32\mathrm{i} & -0.23\mathrm{i} & 0.07-0.24\mathrm{i} \\ -0.06-0.33\mathrm{i} & 0.3-0.67\mathrm{i} & -0.35-0.48\mathrm{i} & 0.38-0.5\mathrm{i} \\ -0.23+0.03\mathrm{i} & -0.38+0.13\mathrm{i} & -0.34+0.18\mathrm{i} & -0.71-0.09\mathrm{i} \end{bmatrix}^{\mathrm{T}}$		3x3 multiplications and 3x2 additions for each column of $\tilde{\mathbf{H}}^{+} = \mathbf{L}\mathbf{K}^{-1}\tilde{\mathbf{H}}^{\mathrm{H}}$, plus 3x3 multiplications to compute $\mathbf{L}\mathbf{K}^{-1}$
	(b)	$k_0 := \underset{j\in[0,2]}{\arg\min} \left\| (\mathbf{G})_j \right\|^2 = 0$		4 multiplications and 3 additions to compute $\left\| (\mathbf{G})_j \right\|^2$ for each of 3 non-zero columns
	(c)	$\tilde{x}(0) = \mathrm{demod}[(\mathbf{G})_0^{\mathrm{T}} \tilde{\mathbf{y}}] = \mathrm{demod}[-0.24+0.54\mathrm{i}] = 1\mathrm{i}$		4 multiplications and 3 additions
	(d)	$\tilde{\mathbf{y}} := \tilde{\mathbf{y}} - (\tilde{\mathbf{H}})_0 \tilde{x}(0) = \begin{bmatrix} -0.01-1.65\mathrm{i} \\ -1.27+0.7\mathrm{i} \\ 0.4-1.98\mathrm{i} \\ 2.15+0.09\mathrm{i} \end{bmatrix}$		4 multiplications and 4 additions
$i = 1$	Init.	$\tilde{\mathbf{H}} := \begin{bmatrix} 0 & -1.02+0.4\mathrm{i} & -0.28+0.49\mathrm{i} \\ 0 & 0.91+0.47\mathrm{i} & 0.84-0.85\mathrm{i} \\ 0 & -1.13+0.18\mathrm{i} & -0.75+\mathrm{i} \\ 0 & 0.16-0.65\mathrm{i} & -1.53-0.04\mathrm{i} \end{bmatrix}$		
	(a)	$\mathbf{K} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4.01 & 1.66-3.55\mathrm{i} \\ 0 & 1.66+3.55\mathrm{i} & 5.65 \end{bmatrix}$, $\mathbf{L} := \mathbf{I}$		
		$r = 0$	$\mathbf{k} := [\mathbf{K}]_{0,0}^{-1} (\mathbf{K}^{\mathrm{T}})_0 - (\mathbf{I})_0 = \mathbf{0}$	0 multiplications and 0 additions

			$\mathbf{K} := \mathbf{K} - (\mathbf{K})_0 \mathbf{k}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4.01 & 1.66 - 3.55\mathrm{i} \\ 0 & 1.66 + 3.55\mathrm{i} & 5.65 \end{bmatrix}$	0 multiplications and 0 additions
			$\mathbf{L} := \mathbf{L} - (\mathbf{L})_0 \mathbf{k}^{\mathrm{T}} = \mathbf{I}$	0 multiplications and 0 additions
		$r = 1$	$\mathbf{k} := [\mathbf{K}]_{1,1}^{-1} (\mathbf{K}^{\mathrm{T}})_1 - (\mathbf{I})_1 = \begin{bmatrix} 0 \\ 0 \\ 0.41 - 0.89\mathrm{i} \end{bmatrix}$	1 multiplication and 0 additions
			$\mathbf{K} := \mathbf{K} - (\mathbf{K})_1 \mathbf{k}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4.01 & 0 \\ 0 & 1.66 + 3.55\mathrm{i} & 1.82 \end{bmatrix}$	1 multiplication for the 3rd column; 1 addition for the 3rd column
			$\mathbf{L} := \mathbf{L} - (\mathbf{L})_1 \mathbf{k}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.41 + 0.89\mathrm{i} \\ 0 & 0 & 1 \end{bmatrix}$	0 multiplications and 0 additions
		$r = 2$	$\mathbf{k} := [\mathbf{K}]_{2,2}^{-1} (\mathbf{K}^{\mathrm{T}})_2 - (\mathbf{I})_2 = \begin{bmatrix} 0 \\ 0.91 + 1.95\mathrm{i} \\ 0 \end{bmatrix}$	1 multiplication and 0 additions
			$\mathbf{K} := \mathbf{K} - (\mathbf{K})_2 \mathbf{k}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4.01 & 0 \\ 0 & 0 & 1.82 \end{bmatrix}$	0 multiplications and 0 additions
			$\mathbf{L} := \mathbf{L} - (\mathbf{L})_2 \mathbf{k}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3.11 & -0.41 + 0.89\mathrm{i} \\ 0 & -0.91 - 1.95\mathrm{i} & 1 \end{bmatrix}$	1 multiplication for the 2nd column; 1 addition for the 2nd column
			$\mathbf{G} := (\mathbf{L}\mathbf{K}^{-1}\tilde{\mathbf{H}}^{\mathrm{H}})^{\mathrm{T}}$ $= \begin{bmatrix} 0 & 0 & 0 & 0 \\ -0.49 - 0.34\mathrm{i} & 0.1 - 0.15\mathrm{i} & -0.22 - 0.28\mathrm{i} & 0.45 - 0.25\mathrm{i} \\ -0.12 + 0.32\mathrm{i} & 0.03 + 0.13\mathrm{i} & -0.24 + 0.04\mathrm{i} & -0.56 - 0.2\mathrm{i} \end{bmatrix}^{\mathrm{T}}$	2x2 multiplications and 2x1 additions for each column of $\tilde{\mathbf{H}}^{+} = \mathbf{L}\mathbf{K}^{-1}\tilde{\mathbf{H}}^{\mathrm{H}}$, plus 2x2 multiplications to compute $\mathbf{L}\mathbf{K}^{-1}$

	(b)	$k_1 := \arg\min_{j\in[1,2]} \left\| (\mathbf{G})_j \right\|^2 = 2$	4 multiplications and 3 additions to compute $\left\| (\mathbf{G})_j \right\|^2$ for each of 2 non-zero columns
	(c)	$\tilde{x}(2) = \text{demod}[(\mathbf{G})_2^{\text{T}} \tilde{\mathbf{y}}] = \text{demod}[-0.8 + 0.05\text{i}] = -1$	4 multiplications and 3 additions
	(d)	$\tilde{\mathbf{y}} := \tilde{\mathbf{y}} - (\tilde{\mathbf{H}})_2 \tilde{x}(2) = \begin{bmatrix} -0.29 - 1.16\text{i} \\ -0.43 - 0.15\text{i} \\ -0.35 - 0.98\text{i} \\ 0.62 + 0.05\text{i} \end{bmatrix}$	4 multiplications and 4 additions
$i = 2$	Init.	$\tilde{\mathbf{H}} := \begin{bmatrix} 0 & -1.02 + 0.4\text{i} & 0 \\ 0 & 0.91 + 0.47\text{i} & 0 \\ 0 & -1.13 + 0.18\text{i} & 0 \\ 0 & 0.16 - 0.65\text{i} & 0 \end{bmatrix}$	
	(a)	$\mathbf{K} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4.01 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\mathbf{L} := \mathbf{I}$	
	$r = 0$	$\mathbf{k} := [\mathbf{K}]_{0,0}^{-1} (\mathbf{K}^{\text{T}})_0 - (\mathbf{I})_0 = \mathbf{0}$	0 multiplications and 0 additions
		$\mathbf{K} := \mathbf{K} - (\mathbf{K})_0 \mathbf{k}^{\text{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4.01 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	0 multiplications and 0 additions
		$\mathbf{L} := \mathbf{L} - (\mathbf{L})_0 \mathbf{k}^{\text{T}} = \mathbf{I}$	0 multiplications and 0 additions
	$r = 1$	$\mathbf{k} := [\mathbf{K}]_{1,1}^{-1} (\mathbf{K}^{\text{T}})_1 - (\mathbf{I})_1 = \mathbf{0}$	0 multiplications and 0 additions
		$\mathbf{K} := \mathbf{K} - (\mathbf{K})_1 \mathbf{k}^{\text{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4.01 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	0 multiplications and 0 additions

			$\mathbf{L} := \mathbf{L} - (\mathbf{L})_1 \mathbf{k}^{\mathrm{T}} = \mathbf{I}$	0 multiplications and 0 additions
		$r = 2$	$\mathbf{k} := [\mathbf{K}]_{2,2}^{-1} (\mathbf{K}^{\mathrm{T}})_2 - (\mathbf{I})_2 = \mathbf{0}$	0 multiplications and 0 additions
			$\mathbf{K} := \mathbf{K} - (\mathbf{K})_2 \mathbf{k}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4.01 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	0 multiplications and 0 additions
			$\mathbf{L} := \mathbf{L} - (\mathbf{L})_2 \mathbf{k}^{\mathrm{T}} = \mathbf{I}$	0 multiplications and 0 additions
		$\mathbf{G} := (\mathbf{L}\mathbf{K}^{-1}\tilde{\mathbf{H}}^{\mathrm{H}})^{\mathrm{T}}$ $= \begin{bmatrix} 0 & 0 & 0 & 0 \\ -0.25-0.1\mathrm{i} & 0.23-0.12\mathrm{i} & -0.28-0.04\mathrm{i} & 0.04+0.16\mathrm{i} \\ 0 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$		4 multiplications and 0 additions
	(b)	$k_2 := 1$		0 multiplications and 0 additions
	(c)	$\tilde{x}(1) = \mathrm{demod}[(\mathbf{G})_1^{\mathrm{T}} \tilde{\mathbf{y}}] = \mathrm{demod}[-0.09 + 0.74\mathrm{i}] = 1\mathrm{i}$		4 multiplications and 3 additions
Out.	$\tilde{\mathbf{x}}_{n,m} = \begin{bmatrix} \tilde{x}(0) \\ \tilde{x}(1) \\ \tilde{x}(2) \end{bmatrix} = \begin{bmatrix} 1\mathrm{i} \\ 1\mathrm{i} \\ -1 \end{bmatrix}$			

It can be seen that ZF-VBLAST detector avoids decision error in the 2nd symbol in contrast to its linear counterpart.

By induction, computational complexity per subcarrier for a general case of $N_{\mathrm{rx}} \times N_{\mathrm{tx}}$ system is determined:

Step in *i*th recursion ($i \in [0, N_{\mathrm{tx}} - 2]$)	**Computed variable**	**Multiplicative complexity** $c_{\mathrm{M}}(i)$	**Additive complexity** $c_{\mathrm{A}}(i)$
(a)	**k**	$(N_{\mathrm{tx}} - i - 1)(N_{\mathrm{tx}} - i)$	0

	K	$(N_{tx}-i-1)\sum_{r=0}^{N_{tx}-i-1}(N_{tx}-i-r-1)$ $=(N_{tx}-i-1)^2(N_{tx}-i)/2$	$(N_{tx}-i-1)\sum_{r=0}^{N_{tx}-i-1}(N_{tx}-i-r-1)$ $=(N_{tx}-i-1)^2(N_{tx}-i)/2$
	L	$(N_{tx}-i-1)\sum_{r=0}^{N_{tx}-i-1}r$ $=(N_{tx}-i-1)^2(N_{tx}-i)/2$	$(N_{tx}-i-1)\sum_{r=0}^{N_{tx}-i-1}r$ $=(N_{tx}-i-1)^2(N_{tx}-i)/2$
	G	$(N_{rx}+1)(N_{tx}-i)^2$	$N_{rx}(N_{tx}-i-1)(N_{tx}-i)$
(b)	k_i	$N_{rx}(N_{tx}-i)$	$(N_{rx}-1)(N_{tx}-i)$
(c)	$\tilde{x}(k_i)$	N_{rx}	$N_{rx}-1$
(d)	$\tilde{\mathbf{y}}$	N_{rx}	N_{rx}

Note that the number of multiplications, $(N_{tx}-i-1)(N_{tx}-i)+(N_{tx}-i-1)^2(N_{tx}-i)+(N_{tx}-i)^2=(N_{tx}-i)^3$, and the number of additions, $(N_{tx}-i-1)^2(N_{tx}-i)=(N_{tx}-i)^3-2(N_{tx}-i)^2+(N_{tx}-i)$, required for computation of the inverse of $\tilde{\mathbf{H}}^{H}\tilde{\mathbf{H}}$, is the same as in the standard Gauss-Jordan elimination algorithm applied to an arbitrary non-singular matrix of the size $(N_{tx}-i)\times(N_{tx}-i)$ (equations (B.1) and (B.2)).

Total number of multiplications:

$$C_M=\underbrace{N_{rx}\sum_{n=1}^{N_{tx}}n}_{\mathbf{A}}+\underbrace{\sum_{i=0}^{N_{tx}-2}c_M(i)}_{i\in[0,N_{tx}-2]}+\underbrace{2N_{rx}}_{i=N_{tx}-1}$$
$$=\left(\frac{1}{3}N_{rx}+\frac{1}{4}N_{tx}+\frac{2}{3}\right)N_{tx}^3+\frac{1}{4}(6N_{rx}-3)N_{tx}^2+\frac{1}{6}(19N_{rx}+5)N_{tx}-(2N_{rx}+1). \quad \text{(B.7)}$$

Total number of additions:

$$C_A=\underbrace{(N_{rx}-1)\sum_{n=1}^{N_{tx}}n}_{\mathbf{A}}+\underbrace{\sum_{i=0}^{N_{tx}-2}c_A(i)}_{i\in[0,N_{tx}-2]}+\underbrace{N_{rx}-1}_{i=N_{tx}-1}$$
$$=\left(\frac{1}{3}N_{rx}+\frac{1}{4}N_{tx}-\frac{1}{6}\right)N_{tx}^3+\left(N_{rx}-\frac{5}{4}\right)N_{tx}^2+\frac{1}{6}(16N_{rx}-11)N_{tx}-(2N_{rx}-1). \quad \text{(B.8)}$$

If the system is square, i.e. $N_{rx}=N_{tx}$, then the complexity equations are reduced to:

$$C_M=\frac{7}{12}N_{tx}^4+\frac{13}{6}N_{tx}^3+\frac{29}{12}N_{tx}^2-\frac{7}{6}N_{tx}-1, \quad \text{(B.9)}$$

$$C_{\mathrm{A}} = \frac{7}{12}N_{\mathrm{tx}}^4 + \frac{5}{6}N_{\mathrm{tx}}^3 + \frac{17}{12}N_{\mathrm{tx}}^2 - \frac{23}{6}N_{\mathrm{tx}} + 1 \ . \tag{B.10}$$

One can also see that the complexity of determining the optimal decoding order in step (b) ($N_{\mathrm{rx}}\sum_{i=0}^{N_{\mathrm{tx}}-2}(N_{\mathrm{tx}}-i) = \frac{1}{2}N_{\mathrm{rx}}N_{\mathrm{tx}}^2 + \frac{1}{2}N_{\mathrm{rx}}N_{\mathrm{tx}} - N_{\mathrm{rx}}$ multiplications and $(N_{\mathrm{rx}}-1)\sum_{i=0}^{N_{\mathrm{tx}}-2}(N_{\mathrm{tx}}-i) = \frac{1}{2}N_{\mathrm{rx}}N_{\mathrm{tx}}^2 + \frac{1}{2}(N_{\mathrm{rx}}-N_{\mathrm{tx}})N_{\mathrm{tx}} - \frac{1}{2}N_{\mathrm{tx}} - (N_{\mathrm{rx}}-1)$ additions) is by an order smaller than the total computational complexity described by (B.7) and (B.8). Thus, the pseudoinverse computation in step (a) determines the order of complexity. It should be noted that for the MMSE-VBLAST detector the inverted matrix $(\tilde{\mathbf{H}}^{\mathrm{H}}\tilde{\mathbf{H}})^{-1}$ has the same size ($N_{\mathrm{tx}} \times N_{\mathrm{tx}}$) as for the ZF-VBLAST case. Thus, the computational effort for both schemes is of the same order, $O[(N_{\mathrm{rx}} + N_{\mathrm{tx}})N_{\mathrm{tx}}^3]$, so we do not derive here the complexity expressions specific to the MMSE-VBLAST.

B.3: Example of QRD detection

The QR decomposition (QRD) applied to the matrix $\tilde{\mathbf{H}}$, having columns sorted in the order of the average noise variance at the detector output, serves a low-complexity alternative to the optimal V-BLAST described in Tab.2.1. Tab.2.3 summarises the sorted QRD (SQRD) algorithm, which is based on the Gram-Schmidt orthogonalisation (GSO) performed in the prescribed sorting order.

To illustrate ZF-weighted SQRD operation, we use the same example as in Appendix B.2 ($N_{\mathrm{tx}} = 3$, $N_{\mathrm{rx}} = 4$ and $\mathbf{x}_{n,m} = [1\mathrm{i} \quad 1\mathrm{i} \quad -1]^{\mathrm{T}}$), where it has also been established that the optimal detection order is {0, 2, 1}. The MIMO transform matrix and the received symbols vector are

$$\mathbf{H}_{n,m} = \begin{bmatrix} 0.89-1.13\mathrm{i} & -1.02+0.4\mathrm{i} & -0.28+0.49\mathrm{i} \\ 0.58+0.5\mathrm{i} & 0.91+0.47\mathrm{i} & 0.84-0.85\mathrm{i} \\ -0.01-0.11\mathrm{i} & -1.13+0.18\mathrm{i} & -0.75+\mathrm{i} \\ -0.57+0.37\mathrm{i} & 0.16-0.65\mathrm{i} & -1.53-0.04\mathrm{i} \end{bmatrix} \text{ and } \mathbf{y}_{n,m} = \begin{bmatrix} 1.12-0.76\mathrm{i} \\ -1.77+1.28\mathrm{i} \\ 0.51-1.99\mathrm{i} \\ 1.78-0.48\mathrm{i} \end{bmatrix}.$$

	Operation record	**Computational load**
Init.	$\tilde{\mathbf{H}} := \begin{bmatrix} 0.89-1.13\mathrm{i} & -1.02+0.4\mathrm{i} & -0.28+0.49\mathrm{i} \\ 0.58+0.5\mathrm{i} & 0.91+0.47\mathrm{i} & 0.84-0.85\mathrm{i} \\ -0.01-0.11\mathrm{i} & -1.13+0.18\mathrm{i} & -0.75+\mathrm{i} \\ -0.57+0.37\mathrm{i} & 0.16-0.65\mathrm{i} & -1.53-0.04\mathrm{i} \end{bmatrix}$	

	$k_0 := 0,\ k_1 := 2,\ k_2 := 1$		
$i = 2$	Init.	$\mathbf{Q} := \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
	(a)	$(\mathbf{P})_1 := \mathbf{Q}^{\mathrm{H}}(\tilde{\mathbf{H}})_1 = \mathbf{0}$	0 multiplications and 0 additions
	(b)	$\mathbf{q} := (\tilde{\mathbf{H}})_1 - \mathbf{Q}(\mathbf{P})_1 = \begin{bmatrix} -1.02 + 0.4\mathrm{i} \\ 0.91 + 0.47\mathrm{i} \\ -1.13 + 0.18\mathrm{i} \\ 0.16 - 0.65\mathrm{i} \end{bmatrix}$	0 multiplications and 0 additions
	(c)	$[\mathbf{P}]_{1,1} := \lVert \mathbf{q} \rVert = 2$	4 multiplications, 3 additions and a square root operation (equivalent to s multiplications)
	(d)	$(\mathbf{Q})_1 := [\mathbf{P}]_{1,1}^{-1}\mathbf{q} = \begin{bmatrix} -0.51 + 0.2\mathrm{i} \\ 0.45 + 0.23\mathrm{i} \\ -0.56 + 0.09\mathrm{i} \\ 0.08 - 0.32\mathrm{i} \end{bmatrix}$	4 multiplications and 0 additions
$i = 1$	Init.	$\mathbf{Q} := \begin{bmatrix} 0 & -0.51 + 0.2\mathrm{i} & 0 \\ 0 & 0.45 + 0.23\mathrm{i} & 0 \\ 0 & -0.56 + 0.09\mathrm{i} & 0 \\ 0 & 0.08 - 0.32\mathrm{i} & 0 \end{bmatrix}$	
	(a)	$(\mathbf{P})_2 := \mathbf{Q}^{\mathrm{H}}(\tilde{\mathbf{H}})_2 = \begin{bmatrix} 0 \\ 0.83 - 1.77\mathrm{i} \\ 0 \end{bmatrix}$	4 multiplications and 3 additions
	(b)	$\mathbf{q} := (\tilde{\mathbf{H}})_2 - \mathbf{Q}(\mathbf{P})_2 = \begin{bmatrix} -0.21 - 0.58\mathrm{i} \\ 0.05 - 0.24\mathrm{i} \\ -0.44 - 0.08\mathrm{i} \\ -1.02 + 0.37\mathrm{i} \end{bmatrix}$	4 multiplications and 4 additions

	(c)	$[\mathbf{P}]_{2,2} := \lVert \mathbf{q} \rVert = 1.35$	4 multiplications, 3 additions and a square root operation
	(d)	$(\mathbf{Q})_2 := [\mathbf{P}]_{2,2}^{-1}\mathbf{q} = \begin{bmatrix} -0.16-0.43\mathrm{i} \\ 0.04-0.18\mathrm{i} \\ -0.33-0.06\mathrm{i} \\ -0.76+0.27\mathrm{i} \end{bmatrix}$	4 multiplications and 0 additions
$i=0$	Init.	$\mathbf{Q} := \begin{bmatrix} 0 & -0.51+0.2\mathrm{i} & -0.16-0.43\mathrm{i} \\ 0 & 0.45+0.23\mathrm{i} & 0.04-0.18\mathrm{i} \\ 0 & -0.56+0.09\mathrm{i} & -0.33-0.06\mathrm{i} \\ 0 & 0.08-0.32\mathrm{i} & -0.76+0.27\mathrm{i} \end{bmatrix}$	
	(a)	$(\mathbf{P})_0 := \mathbf{Q}^{\mathrm{H}}(\tilde{\mathbf{H}})_0 = \begin{bmatrix} 0 \\ -0.47+0.4\mathrm{i} \\ 0.82+0.59\mathrm{i} \end{bmatrix}$	4x2 multiplications and 3x2 additions
	(b)	$\mathbf{q} := (\tilde{\mathbf{H}})_0 - \mathbf{Q}(\mathbf{P})_0 = \begin{bmatrix} 0.61-0.39\mathrm{i} \\ 0.75+0.55\mathrm{i} \\ 0.4\mathrm{i} \\ 0.12+0.41\mathrm{i} \end{bmatrix}$	4x2 multiplications and 4+4x1 additions
	(c)	$[\mathbf{P}]_{0,0} := \lVert \mathbf{q} \rVert = 1.32$	4 multiplications, 3 additions and a square root operation
	(d)	$(\mathbf{Q})_0 := [\mathbf{P}]_{0,0}^{-1}\mathbf{q} = \begin{bmatrix} 0.46-0.3\mathrm{i} \\ 0.57+0.42\mathrm{i} \\ 0.3\mathrm{i} \\ 0.09+0.31\mathrm{i} \end{bmatrix}$	4 multiplications and 0 additions
$\mathbf{Q} := \begin{bmatrix} 0.46-0.3\mathrm{i} & -0.51+0.2\mathrm{i} & -0.16-0.43\mathrm{i} \\ 0.57+0.42\mathrm{i} & 0.45+0.23\mathrm{i} & 0.04-0.18\mathrm{i} \\ 0.3\mathrm{i} & -0.56+0.09\mathrm{i} & -0.33-0.06\mathrm{i} \\ 0.09+0.31\mathrm{i} & 0.08-0.32\mathrm{i} & -0.76+0.27\mathrm{i} \end{bmatrix}$			
$\tilde{\mathbf{y}} := \mathbf{Q}^{\mathrm{H}}\mathbf{y}_{n,m} = \begin{bmatrix} -0.32+0.71\mathrm{i} \\ -1.4+2.78\mathrm{i} \\ -1.67+0.89\mathrm{i} \end{bmatrix}$			3x4 multiplications and 3x3 additions

$i=0$	(e)	$\tilde{x}(0) := \text{demod}\left[[\mathbf{P}]_{0,0}^{-1}[\tilde{\mathbf{y}}]_0\right] = \text{demod}[-0.24+0.54\text{i}] = 1\text{i}$	1 multiplication and 0 additions
	(f)	$\tilde{\mathbf{y}} := \tilde{\mathbf{y}} - (\mathbf{P})_0 \tilde{x}(0) = \begin{bmatrix} -0.32-0.61\text{i} \\ -1+3.25\text{i} \\ -1.08+0.07\text{i} \end{bmatrix}$	3 multiplications and 3 additions
$i=1$	(e)	$\tilde{x}(2) := \text{demod}\left[[\mathbf{P}]_{2,2}^{-1}[\tilde{\mathbf{y}}]_2\right] = \text{demod}[-0.8+0.05\text{i}] = -1$	1 multiplication and 0 additions
	(f)	$\tilde{\mathbf{y}} := \tilde{\mathbf{y}} - (\mathbf{P})_0 \tilde{x}(2) = \begin{bmatrix} -0.32-0.61\text{i} \\ -0.17+1.47\text{i} \\ 0.27+0.07\text{i} \end{bmatrix}$	3 multiplications and 3 additions
$i=2$	(e)	$\tilde{x}(1) := \text{demod}\left[[\mathbf{P}]_{1,1}^{-1}[\tilde{\mathbf{y}}]_1\right] = \text{demod}[-0.09+0.74\text{i}] = 1\text{i}$	1 multiplication and 0 additions
Out.	$\tilde{\mathbf{x}}_{n,m} = \begin{bmatrix} \tilde{x}(0) \\ \tilde{x}(1) \\ \tilde{x}(2) \end{bmatrix} = \begin{bmatrix} 1\text{i} \\ 1\text{i} \\ -1 \end{bmatrix}$		

Similar to ZF-VBLAST, ZF-SQRD detector avoids decision error in the 2nd symbol in contrast to the linear ZF, despite the ordering does not get updated in the detection process.

Using the observations from the aforementioned example, computational complexity per subcarrier for ZF-SQRD is determined in the sequel for a general case of $N_{\text{rx}} \times N_{\text{tx}}$ system. It is quite simple to show that the complexity for the MMSE-SQRD algorithm variant is of the same order as that of ZF-SQRD. Thus, we derive the explicit complexity expressions only for ZF-SQRD.

Step in *i*th recursion ($i \in [0, N_{\text{tx}}-1]$)	**Computed variable**	**Multiplicative complexity** $c_{\text{M}}(i)$	**Additive complexity** $c_{\text{A}}(i)$
(a)	$(\mathbf{P})_{k_i}$	$N_{\text{rx}} i$	$(N_{\text{rx}}-1)i$
(b)	$\mathbf{q}$	$N_{\text{rx}} i$	$N_{\text{rx}} i$
(c)	$[\mathbf{P}]_{k_i,k_i}$	$N_{\text{rx}} + s$	$N_{\text{rx}} - 1$

(d)	$(\mathbf{Q})_{k_i}$	N_{rx}	0
(e)	$\tilde{x}(k_i)$	1	0
(f)	$\tilde{\mathbf{y}}$ ($i \in [0, N_{\text{tx}} - 2]$)	N_{tx}	N_{tx}

It should be noted that GSO (steps (a)-(d)) requires $\sum_{i=0}^{N_{\text{tx}}-1}(2N_{\text{rx}}i + 2N_{\text{rx}} + s) = N_{\text{rx}}N_{\text{tx}}(N_{\text{tx}} - 1) + (2N_{\text{rx}} + s)N_{\text{tx}} = N_{\text{rx}}N_{\text{tx}}^2 + (N_{\text{rx}} + s)N_{\text{tx}}$ multiplications and $\sum_{i=0}^{N_{\text{tx}}-1}(2N_{\text{rx}}i - i + N_{\text{rx}} - 1) = N_{\text{rx}}N_{\text{tx}}(N_{\text{tx}} - 1) - \frac{1}{2}N_{\text{tx}}(N_{\text{tx}} - 1) + (N_{\text{rx}} - 1)N_{\text{tx}} = \left(N_{\text{rx}} - \frac{1}{2}\right)N_{\text{tx}}^2 - \frac{1}{2}N_{\text{tx}}$ additions.

Total number of multiplications:

$$\begin{aligned} C_{\text{M}} &= N_{\text{rx}}N_{\text{tx}} - N_{\text{tx}} + \sum_{i=0}^{N_{\text{tx}}-1} c_{\text{M}}(i) \\ &= (N_{\text{rx}} - 1)N_{\text{tx}} + N_{\text{rx}}N_{\text{tx}}(N_{\text{tx}} - 1) + (2N_{\text{rx}} + s)N_{\text{tx}} + N_{\text{tx}} + N_{\text{tx}}^2 \\ &= (N_{\text{rx}} + 1)N_{\text{tx}}^2 + (2N_{\text{rx}} + s)N_{\text{tx}} \end{aligned} \quad . \tag{B.11}$$

Total number of additions:

$$\begin{aligned} C_{\text{A}} &= (N_{\text{rx}} - 1)N_{\text{tx}} - N_{\text{tx}} + \sum_{i=0}^{N_{\text{tx}}-1} c_{\text{A}}(i) \\ &= (N_{\text{rx}} - 2)N_{\text{tx}} + N_{\text{rx}}N_{\text{tx}}(N_{\text{tx}} - 1) - \frac{1}{2}N_{\text{tx}}(N_{\text{tx}} - 1) + (N_{\text{rx}} - 1)N_{\text{tx}} + N_{\text{tx}}^2 \\ &= \left(N_{\text{rx}} + \frac{1}{2}\right)N_{\text{tx}}^2 + \left(N_{\text{rx}} - \frac{5}{2}\right)N_{\text{tx}} \end{aligned} \quad . \tag{B.12}$$

If the system is square, i.e. $N_{\text{rx}} = N_{\text{tx}}$, then the complexity expressions are reduced to

$$C_{\text{M}} = N_{\text{tx}}^3 + 3N_{\text{tx}}^2 + sN_{\text{tx}}, \tag{B.13}$$

$$C_{\text{A}} = N_{\text{tx}}^3 + \frac{3}{2}N_{\text{tx}}^2 - \frac{5}{2}N_{\text{tx}}. \tag{B.14}$$

Complexity formulas (B.11) and (B.12) do not take into account computation of the ordering. Previous research works [129] point out that the once-off calculation of all the SNR priorities, performed in the first ($i = 0$) recursion of the classic V-BLAST algorithm (step (b) in Tab.2.1), provides a quite accurate ordering estimate. Thus, $\tilde{\mathbf{H}}^{+} = \mathbf{H}_{n,m}^{+}$ is computed only once to determine the suboptimal detection sequence. Based on the low complexity SQRD (Tab.2.3), $\tilde{\mathbf{H}} = \mathbf{Q}\mathbf{P}$, one can express $\tilde{\mathbf{H}}^{+} = (\tilde{\mathbf{H}}^{\text{H}}\tilde{\mathbf{H}})^{-1}\tilde{\mathbf{H}}^{\text{H}} = \mathbf{P}^{-1}\mathbf{Q}^{\text{H}}$. Unitarity of $\mathbf{Q}$ implies that $\left\|(\tilde{\mathbf{H}}^{+\text{T}})_j\right\| = \left\|(\mathbf{P}^{-\text{T}})_j\right\|$, where $j \in [0, N_{\text{tx}} - 1]$. Hence computation of the inverse is greatly simplified due to the permuted triangular shape of $\mathbf{P}$, requiring only $\frac{1}{6}N_{\text{tx}}^3 + \frac{1}{2}N_{\text{tx}}^2 + \frac{1}{3}N_{\text{tx}}$ multiplications and $\frac{1}{6}N_{\text{tx}}^3 - \frac{1}{2}N_{\text{tx}}^2 + \frac{1}{3}N_{\text{tx}}$

additions, i.e. $O(N_{\text{tx}}^3)$ mathematical operations. Subsequent calculation of the square norms of $\mathbf{P}^{-\text{T}}$ columns involves $\sum_{j=0}^{N_{\text{tx}}-1}(j+1)=\frac{1}{2}N_{\text{tx}}(N_{\text{tx}}+1)$ multiplications and $\sum_{j=0}^{N_{\text{tx}}-1}j=\frac{1}{2}N_{\text{tx}}(N_{\text{tx}}-1)$ additions, i.e. $O(N_{\text{tx}}^2)$ operations.

With the necessity to determine ordering of the detection, the algorithm consists of two main stages:

1) primary GSO to compute the standard QR decomposition $\tilde{\mathbf{H}}=\mathbf{Q_1}\mathbf{P_1}$, followed by the inversion of $\mathbf{P_1}$ using the Gauss-Jordan elimination to compute ordering (Tab.2.4);
2) secondary GSO to compute SQRD $\tilde{\mathbf{H}}=\mathbf{Q_2}\mathbf{P_2}$ (Tab.2.3), followed by the ordered detection procedure based on $\mathbf{Q_2}$ and $\mathbf{P_2}$. It should be noted that in general $\mathbf{Q_1}\neq\mathbf{Q_2}$ and $\mathbf{P_1}\neq\mathbf{P_2}$.

Thus, the overall computational complexity in case of the pre-detection ordering constitutes:

$$\begin{aligned}C_{\text{M}}&=N_{\text{rx}}N_{\text{tx}}^2+(N_{\text{rx}}+s)N_{\text{tx}}+\frac{1}{6}N_{\text{tx}}^3+\frac{1}{2}N_{\text{tx}}^2+\frac{1}{3}N_{\text{tx}}+\frac{1}{2}N_{\text{tx}}(N_{\text{tx}}+1)+(N_{\text{rx}}+1)N_{\text{tx}}^2+(2N_{\text{rx}}+s)N_{\text{tx}}\\&=\left(2N_{\text{rx}}+\frac{1}{6}N_{\text{tx}}+2\right)N_{\text{tx}}^2+\left(3N_{\text{rx}}+2s+\frac{5}{6}\right)N_{\text{tx}}\end{aligned},\quad\text{(B.15)}$$

$$\begin{aligned}C_{\text{A}}&=\left(N_{\text{rx}}-\frac{1}{2}\right)N_{\text{tx}}^2-\frac{1}{2}N_{\text{tx}}+\frac{1}{6}N_{\text{tx}}^3-\frac{1}{2}N_{\text{tx}}^2+\frac{1}{3}N_{\text{tx}}+\frac{1}{2}N_{\text{tx}}(N_{\text{tx}}-1)+\left(N_{\text{rx}}+\frac{1}{2}\right)N_{\text{tx}}^2+\left(N_{\text{rx}}-\frac{5}{2}\right)N_{\text{tx}}\\&=\left(2N_{\text{rx}}+\frac{1}{6}N_{\text{tx}}+\frac{1}{2}\right)N_{\text{tx}}^2+\left(N_{\text{rx}}-\frac{19}{6}\right)N_{\text{tx}}\end{aligned}.\quad\text{(B.16)}$$

For the square system these complexity expressions are reduced to:

$$C_{\text{M}}=\frac{13}{6}N_{\text{tx}}^3+5N_{\text{tx}}^2+\left(2s+\frac{5}{6}\right)N_{\text{tx}},\quad\text{(B.17)}$$

$$C_{\text{A}}=\frac{13}{6}N_{\text{tx}}^3+\frac{3}{2}N_{\text{tx}}^2-\frac{19}{6}N_{\text{tx}}.\quad\text{(B.18)}$$

If the channel variation in between the subsequent blocks is insignificant, then there is no need for the primary GSO as the ordering remains unchanged. In such a case it is sufficient to update SNR weights after detection; hence this approach will be termed post-detection ordering. The corresponding system complexity is subject to approximately two-fold reduction:

$$\begin{aligned}C_{\text{M}}&=\left(2N_{\text{rx}}+\frac{1}{6}N_{\text{tx}}+2\right)N_{\text{tx}}^2+\left(3N_{\text{rx}}+2s+\frac{5}{6}\right)N_{\text{tx}}-N_{\text{rx}}N_{\text{tx}}^2-(N_{\text{rx}}+s)N_{\text{tx}}\\&=\left(N_{\text{rx}}+\frac{1}{6}N_{\text{tx}}+2\right)N_{\text{tx}}^2+\left(2N_{\text{rx}}+s+\frac{5}{6}\right)N_{\text{tx}}\end{aligned},\quad\text{(B.19)}$$

$$\begin{aligned}C_{\text{A}}&=\left(2N_{\text{rx}}+\frac{1}{6}N_{\text{tx}}+\frac{1}{2}\right)N_{\text{tx}}^2+\left(N_{\text{rx}}-\frac{19}{6}\right)N_{\text{tx}}-\left(N_{\text{rx}}-\frac{1}{2}\right)N_{\text{tx}}^2+\frac{1}{2}N_{\text{tx}}\\&=\left(N_{\text{rx}}+\frac{1}{6}N_{\text{tx}}+1\right)N_{\text{tx}}^2+\left(N_{\text{rx}}-\frac{8}{3}\right)N_{\text{tx}}\end{aligned}.\quad\text{(B.20)}$$

For the square system it is described as

$$C_{\text{M}}=\frac{7}{6}N_{\text{tx}}^3+4N_{\text{tx}}^2+\left(s+\frac{5}{6}\right)N_{\text{tx}},\quad\text{(B.21)}$$

$$C_{\mathrm{A}} = \frac{7}{6}N_{\mathrm{tx}}^{3} + 2N_{\mathrm{tx}}^{2} - \frac{8}{3}N_{\mathrm{tx}}\,. \tag{B.22}$$

Appendix C

C.1: Derivation of CLS estimator

For brief notation we make a substitution of the observation matrix $\mathbf{G} = \mathbf{X}^{\mathrm{p}}\mathbf{C}\mathbf{F}\mathbf{B}$ in (3.6). Then we can put down:

$$J(\underline{\mathbf{h}}) = \left(\mathbf{y}^{\mathrm{p}} - \mathbf{G}\underline{\mathbf{h}}\right)^{\mathrm{H}}\left(\mathbf{y}^{\mathrm{p}} - \mathbf{G}\underline{\mathbf{h}}\right) = \mathbf{y}^{\mathrm{p}\,\mathrm{H}}\mathbf{y}^{\mathrm{p}} - \underline{\mathbf{h}}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{y}^{\mathrm{p}} - \mathbf{y}^{\mathrm{p}\,\mathrm{H}}\mathbf{G}\underline{\mathbf{h}} + \underline{\mathbf{h}}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{G}\underline{\mathbf{h}}. \tag{C.1}$$

The CLS estimate of $\underline{\mathbf{h}}$ is obtained by minimisation of (C.1) with respect to the complex-valued vector $\underline{\mathbf{h}}$, i.e. $\underline{\tilde{\mathbf{h}}} = \arg\min_{\underline{\mathbf{h}}} J(\underline{\mathbf{h}})$, that necessitates calculation of the gradient:

$$\nabla_{\underline{\mathbf{h}}} J(\underline{\mathbf{h}}) = \partial J(\underline{\mathbf{h}})/\partial\Re\underline{\mathbf{h}} + \mathrm{j}\partial J(\underline{\mathbf{h}})/\partial\Im\underline{\mathbf{h}}, \tag{C.2}$$

where $\underline{\mathbf{h}} = \Re\underline{\mathbf{h}} + \mathrm{j}\,\Im\underline{\mathbf{h}}$, and (C.1) is correspondingly expanded to

$$\begin{aligned} J(\underline{\mathbf{h}}) &= \mathbf{y}^{\mathrm{p}\,\mathrm{H}}\mathbf{y}^{\mathrm{p}} - \underline{\mathbf{h}}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{y}^{\mathrm{p}} - \mathbf{y}^{\mathrm{p}\,\mathrm{H}}\mathbf{G}\underline{\mathbf{h}} + \underline{\mathbf{h}}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\mathbf{G}\underline{\mathbf{h}} \\ &= \mathbf{y}^{\mathrm{p}\,\mathrm{H}}\mathbf{y}^{\mathrm{p}} - \Re\underline{\mathbf{h}}^{\mathrm{T}}\mathbf{G}^{\mathrm{H}}\mathbf{y}^{\mathrm{p}} + \mathrm{j}\Im\underline{\mathbf{h}}^{\mathrm{T}}\mathbf{G}^{\mathrm{H}}\mathbf{y}^{\mathrm{p}} - \mathbf{y}^{\mathrm{p}\,\mathrm{H}}\mathbf{G}\Re\underline{\mathbf{h}} - \mathrm{j}\mathbf{y}^{\mathrm{p}\,\mathrm{H}}\mathbf{G}\Im\underline{\mathbf{h}} \\ &\quad + \Re\underline{\mathbf{h}}^{\mathrm{T}}\mathbf{G}^{\mathrm{H}}\mathbf{G}\Re\underline{\mathbf{h}} + \Im\underline{\mathbf{h}}^{\mathrm{T}}\mathbf{G}^{\mathrm{H}}\mathbf{G}\Im\underline{\mathbf{h}} - \mathrm{j}\Im\underline{\mathbf{h}}^{\mathrm{T}}\mathbf{G}^{\mathrm{H}}\mathbf{G}\Re\underline{\mathbf{h}} + \mathrm{j}\Re\underline{\mathbf{h}}^{\mathrm{T}}\mathbf{G}^{\mathrm{H}}\mathbf{G}\Im\underline{\mathbf{h}}. \end{aligned} \tag{C.3}$$

In the following derivation, we use expressions of the first and second-order derivatives of a scalar function (any summand in (C.3)) with respect to a real-valued vector [138]:

$$\partial\mathbf{x}^{\mathrm{T}}\mathbf{a}/\partial\mathbf{x} = \partial\mathbf{a}^{\mathrm{T}}\mathbf{x}/\partial\mathbf{x} = \mathbf{a}, \tag{C.4}$$

$$\partial\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}/\partial\mathbf{x} = \left(\mathbf{A} + \mathbf{A}^{\mathrm{T}}\right)\mathbf{x}. \tag{C.5}$$

Hence

$$\partial J(\underline{\mathbf{h}})/\partial\Re\underline{\mathbf{h}} = -\mathbf{G}^{\mathrm{H}}\mathbf{y}^{\mathrm{p}} - \left(\mathbf{y}^{\mathrm{p}\,\mathrm{H}}\mathbf{G}\right)^{\mathrm{T}} + \mathbf{G}^{\mathrm{H}}\mathbf{G}\Re\underline{\mathbf{h}} + \left(\mathbf{G}^{\mathrm{H}}\mathbf{G}\right)^{\mathrm{T}}\Re\underline{\mathbf{h}} - \mathrm{j}\left(\Im\underline{\mathbf{h}}^{\mathrm{T}}\mathbf{G}^{\mathrm{H}}\mathbf{G}\right)^{\mathrm{T}} + \mathrm{j}\mathbf{G}^{\mathrm{H}}\mathbf{G}\Im\underline{\mathbf{h}}, \tag{C.6}$$

and

$$\partial J(\underline{\mathbf{h}})/\partial\Im\underline{\mathbf{h}} = \mathrm{j}\mathbf{G}^{\mathrm{H}}\mathbf{y}^{\mathrm{p}} - \mathrm{j}\left(\mathbf{y}^{\mathrm{p}\,\mathrm{H}}\mathbf{G}\right)^{\mathrm{T}} + \mathbf{G}^{\mathrm{H}}\mathbf{G}\Im\underline{\mathbf{h}} + \left(\mathbf{G}^{\mathrm{H}}\mathbf{G}\right)^{\mathrm{T}}\Im\underline{\mathbf{h}} - \mathrm{j}\mathbf{G}^{\mathrm{H}}\mathbf{G}\Re\underline{\mathbf{h}} + \mathrm{j}\left(\Re\underline{\mathbf{h}}^{\mathrm{T}}\mathbf{G}^{\mathrm{H}}\mathbf{G}\right)^{\mathrm{T}}. \tag{C.7}$$

Substituting (C.6) and (C.7) into (C.2), we obtain:

$$\nabla_{\underline{\mathbf{h}}} J(\underline{\mathbf{h}}) = \partial J(\underline{\mathbf{h}})/\partial\Re\underline{\mathbf{h}} + \mathrm{j}\partial J(\underline{\mathbf{h}})/\partial\Im\underline{\mathbf{h}} = -2\mathbf{G}^{\mathrm{H}}\mathbf{y}^{\mathrm{p}} + 2\mathbf{G}^{\mathrm{H}}\mathbf{G}\left(\Re\underline{\mathbf{h}} + \mathrm{j}\Im\underline{\mathbf{h}}\right) = 2\left(\mathbf{G}^{\mathrm{H}}\mathbf{G}\underline{\mathbf{h}} - \mathbf{G}^{\mathrm{H}}\mathbf{y}^{\mathrm{p}}\right). \tag{C.8}$$

Setting $\nabla_{\underline{\mathbf{h}}} J(\underline{\mathbf{h}}) = \mathbf{0}_{L\times 1}$ yields the CLS estimate of $\underline{\mathbf{h}}$:

$$\underline{\tilde{\mathbf{h}}} = \left(\mathbf{G}^{\mathrm{H}}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{H}}\mathbf{y}^{\mathrm{p}} = \mathbf{G}^{+}\mathbf{y}^{\mathrm{p}} = \sigma_{\mathrm{p}}^{-2}\mathbf{S}\,\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathrm{p}\,\mathrm{H}}\mathbf{y}^{\mathrm{p}}, \tag{C.9}$$

where $\mathbf{S} = \left(\mathbf{G}^{\mathrm{H}}\mathbf{G}\right)^{-1} = \left(\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\,\mathbf{F}\,\mathbf{B}\right)^{-1}$.

Assuming that the elements of the observation vector, $\mathbf{y}^{\mathrm{p}}$, are affected by WGN with zero mean, the derived CLS estimator (C.9) becomes the best linear unbiased estimator (BLUE) [39], for which $\mathrm{E}[\underline{\tilde{\mathbf{h}}}] = \underline{\mathbf{h}}$.

C.2: Design properties of the comb pilot pattern

Let the eigen-decomposed matrix $\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\,\mathbf{F}$, constituting $\mathbf{S} = \left(\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\,\mathbf{F}\,\mathbf{B}\right)^{-1}$, be expressed in the vector combination form [103] as

$$\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\,\mathbf{F} = \sum_{p_k,\,k\in[0,P-1]} (\mathbf{F})_{p_k}\,(\mathbf{F})_{p_k}^{\mathrm{H}}\ . \tag{C.10}$$

From here, using the definition of $\mathbf{F}$ ($[\mathbf{F}]_{m,n} = \exp(-\mathrm{j}2\pi mn/N)$, $m,n\in[0,N-1]$) from (2.1), the (m,n)th element of $\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\,\mathbf{F}$ is found as

$$[\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\,\mathbf{F}]_{m,n} = \sum_{p_k,\,k\in[0,P-1]} [\mathbf{F}]_{m,p_k}\,[\mathbf{F}]_{n,p_k}^{\mathrm{H}} = \sum_{p_k,\,k\in[0,P-1]} \exp[\mathrm{j}2\pi(n-m)\,p_k/N]\,. \tag{C.11}$$

When the pilot subcarriers are equispaced, $N \bmod P = 0$ and $p_k = kN/P + \nu$, where $0 \le \nu \le N/P - 1$. Hence

$$\begin{aligned}[\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\,\mathbf{F}]_{m,n} &= \exp[\mathrm{j}2\pi\nu(n-m)/N]\sum_{k=0}^{P-1}\exp[\mathrm{j}2\pi k(n-m)/P] \\ &= \exp\left(\mathrm{j}\frac{2\pi\nu(n-m)}{N} + \mathrm{j}\frac{\pi(P-1)(n-m)}{P}\right)\frac{\sin[\pi(n-m)]}{\sin[\pi(n-m)/P]} \\ &= \frac{\sin[\pi(n-m)]}{\sin[\pi(n-m)/P]}\exp\left(\mathrm{j}\frac{\pi(n-m)(2P\nu + PN - N)}{PN}\right)\end{aligned}, \tag{C.12}$$

where the sum of the exponential series has been found according to the formula [103]

$$\sum_{k=0}^{l}\mathrm{e}^{\mathrm{j}xl} = \frac{1-\mathrm{e}^{\mathrm{j}(l+1)x}}{1-\mathrm{e}^{\mathrm{j}x}} = \frac{\sin\dfrac{(l+1)x}{2}}{\sin\dfrac{x}{2}}\mathrm{e}^{\mathrm{j}\frac{lx}{2}}. \tag{C.13}$$

The product $\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\,\mathbf{F}\,\mathbf{B} = \mathbf{S}^{-1}$ truncates the last $N-L$ columns and rows of $\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\,\mathbf{F}$, implying $n-m\in[-L+1, L-1]$, where $L\le P$. Taking this constraint into account, one can see from (C.12) that

$$[\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\,\mathbf{F}\,\mathbf{B}]_{m,n} = \begin{cases} P, & m = n \\ 0, & \text{otherwise}\end{cases}. \tag{C.14}$$

Hence $\mathbf{S}^{-1}$ is of the diagonal form, i.e. $\mathbf{S} = P^{-1}\mathbf{I}$.

C.3: Derivation of MMSE estimator

Substituting $\mathbf{y}^{\mathrm{p}} = \mathbf{X}^{\mathrm{p}}\mathbf{C}\mathbf{h} + \mathbf{w}^{\mathrm{p}}$ into (3.18) expands the MSE expression to

$$\begin{aligned} M(\mathbf{Q}) &= \mathrm{E}[(\mathbf{Q}\mathbf{y}^{\mathrm{p}} - \mathbf{h})^{\mathrm{H}}(\mathbf{Q}\mathbf{y}^{\mathrm{p}} - \mathbf{h})] \\ &= \mathrm{E}[(\mathbf{Q}\mathbf{X}^{\mathrm{p}}\mathbf{C}\mathbf{h} + \mathbf{Q}\mathbf{w}^{\mathrm{p}} - \mathbf{h})^{\mathrm{H}}(\mathbf{Q}\mathbf{X}^{\mathrm{p}}\mathbf{C}\mathbf{h} + \mathbf{Q}\mathbf{w}^{\mathrm{p}} - \mathbf{h})] \\ &= \mathrm{E}[(\mathbf{Q}\mathbf{Z}\mathbf{h})^{\mathrm{H}}(\mathbf{Q}\mathbf{Z}\mathbf{h})] + \mathrm{E}[\mathbf{h}^{\mathrm{H}}\mathbf{h}] + \mathrm{E}[(\mathbf{Q}\mathbf{w}^{\mathrm{p}})^{\mathrm{H}}\mathbf{Q}\mathbf{w}^{\mathrm{p}}] - \mathrm{E}[\mathbf{h}^{\mathrm{H}}\mathbf{Q}\mathbf{Z}\mathbf{h}] - \mathrm{E}[\mathbf{h}^{\mathrm{H}}\mathbf{Z}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}}\mathbf{h}] \end{aligned}, \quad \text{(C.15)}$$

where we let $\mathbf{Z} = \mathbf{X}^{\mathrm{p}}\mathbf{C}$ for the brevity of notation.

Using the identities from [102],

$$\mathrm{E}[(\mathbf{A}\mathbf{x})^{\mathrm{H}}\mathbf{A}\mathbf{x}] = \mathrm{tr}(\mathbf{A}\,\mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}]\mathbf{A}^{\mathrm{H}}), \quad \text{(C.16)}$$

$$\mathrm{E}[\mathbf{x}^{\mathrm{H}}\mathbf{B}\mathbf{x}] = \mathrm{tr}(\mathbf{B}\,\mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}]) = \mathrm{tr}(\mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}]\mathbf{B}), \quad \text{(C.17)}$$

one can rewrite (C.15) in the form of

$$\begin{aligned} M(\mathbf{Q}) &= \mathrm{tr}[\mathbf{Q}\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}} + \mathbf{R}_{\mathbf{hh}} + \sigma_{\mathrm{w}}^2\mathbf{Q}\mathbf{Q}^{\mathrm{H}} - \mathbf{Q}\mathbf{Z}\mathbf{R}_{\mathbf{hh}} - \mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}}] \\ &= \mathrm{tr}[\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I})\mathbf{Q}^{\mathrm{H}} + \mathbf{R}_{\mathbf{hh}} - \mathbf{Q}\mathbf{Z}\mathbf{R}_{\mathbf{hh}} - \mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}}\mathbf{Q}^{\mathrm{H}}] \end{aligned}, \quad \text{(C.18)}$$

where $\mathbf{R}_{\mathbf{hh}} = \mathrm{E}[\mathbf{h}\mathbf{h}^{\mathrm{H}}]$ is the CFR correlation matrix and $\mathbf{R}_{\mathbf{w}^{\mathrm{p}}\mathbf{w}^{\mathrm{p}}} = \mathrm{E}[\mathbf{w}^{\mathrm{p}}\mathbf{w}^{\mathrm{p}\mathrm{H}}] = \sigma_{\mathrm{w}}^2\mathbf{I}$ is the WGN correlation matrix.

In general, the elements of $\mathbf{Q}$ are complex-valued. Therefore minimisation of (C.18) with respect to $\mathbf{Q}$ necessitates calculation of the complex gradient matrix

$$\nabla_{\mathbf{Q}} M(\mathbf{Q}) = \partial M(\mathbf{Q})/\partial\Re\mathbf{Q} + \mathrm{j}\partial M(\mathbf{Q})/\partial\Im\mathbf{Q}, \quad \text{(C.19)}$$

where $\mathbf{Q} = \Re\mathbf{Q} + \mathrm{j}\Im\mathbf{Q}$, and (C.18) is correspondingly expanded to:

$$\begin{aligned} M(\mathbf{Q}) &= \mathrm{tr}[\Re\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I})\Re\mathbf{Q}^{\mathrm{T}}] + \mathrm{tr}[\Im\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I})\Im\mathbf{Q}^{\mathrm{T}} + \mathrm{j}\Im\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I})\Re\mathbf{Q}^{\mathrm{T}} \\ &\quad - \mathrm{j}\Re\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I})\Im\mathbf{Q}^{\mathrm{T}}] + \mathrm{tr}[\mathbf{R}_{\mathbf{hh}} - \Re\mathbf{Q}\mathbf{Z}\mathbf{R}_{\mathbf{hh}} - \mathrm{j}\Im\mathbf{Q}\mathbf{Z}\mathbf{R}_{\mathbf{hh}} - \mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}}\Re\mathbf{Q}^{\mathrm{T}} + \mathrm{j}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}}\Im\mathbf{Q}^{\mathrm{T}}]. \end{aligned} \quad \text{(C.20)}$$

In the following derivation, we use expressions of the first and second-order derivatives of trace with respect to a real-valued matrix [138]:

$$\partial\mathrm{tr}(\mathbf{X}\mathbf{A})/\partial\mathbf{X} = \partial\mathrm{tr}(\mathbf{A}\mathbf{X})/\partial\mathbf{X} = \mathbf{A}^{\mathrm{T}}, \quad \text{(C.21)}$$

$$\partial\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{A})/\partial\mathbf{X} = \partial\mathrm{tr}(\mathbf{A}\mathbf{X}^{\mathrm{T}})/\partial\mathbf{X} = \mathbf{A}, \quad \text{(C.22)}$$

$$\partial\mathrm{tr}(\mathbf{X}\mathbf{A}\mathbf{X}^{\mathrm{T}})/\partial\mathbf{X} = \mathbf{X}(\mathbf{A}^{\mathrm{T}} + \mathbf{A}). \quad \text{(C.23)}$$

Hence

$$\begin{aligned} \partial M(\mathbf{Q})/\partial\Re\mathbf{Q} &= \Re\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I})^{\mathrm{T}} + \Re\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I}) + \mathrm{j}\Im\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I}) \\ &\quad - \mathrm{j}\Im\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I})^{\mathrm{T}} - (\mathbf{Z}\mathbf{R}_{\mathbf{hh}})^{\mathrm{T}} - \mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} \end{aligned}, \quad \text{(C.24)}$$

and

$$\partial M(\mathbf{Q})/\partial \Im\mathbf{Q} = \Im\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I})^{\mathrm{T}} + \Im\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I}) + \mathrm{j}\Re\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I})^{\mathrm{T}}$$
$$- \mathrm{j}\Re\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I}) - \mathrm{j}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}})^{\mathrm{T}} + \mathrm{j}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} \quad . \tag{C.25}$$

Substituting (C.24) and (C.25) into (C.19), we obtain:

$$\begin{aligned}\nabla_{\mathbf{Q}} M(\mathbf{Q}) &= \partial M(\mathbf{Q})/\partial \Re\mathbf{Q} + \mathrm{j}\partial M(\mathbf{Q})/\partial \Im\mathbf{Q} \\ &= 2(\Re\mathbf{Q} + \mathrm{j}\Im\mathbf{Q})(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I}) - 2\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} \\ &= 2\mathbf{Q}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I}) - 2\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} \quad .\end{aligned} \tag{C.26}$$

Setting $\nabla_{\mathbf{Q}} M(\mathbf{Q}) = \mathbf{0}_{N\times P}$ yields the expression for the MMSE weighting matrix $\mathbf{Q}$:

$$\begin{aligned}\mathbf{Q} &= \mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}}(\mathbf{Z}\mathbf{R}_{\mathbf{hh}}\mathbf{Z}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I})^{-1} \\ &= \mathbf{R}_{\mathbf{hh}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\mathrm{H}}[\mathbf{X}^{\mathbf{p}}\mathbf{C}\mathbf{R}_{\mathbf{hh}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\mathrm{H}} + \sigma_{\mathrm{w}}^2\mathbf{I}]^{-1} \\ &= \mathbf{R}_{\mathbf{hh}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\mathrm{H}}[\mathbf{X}^{\mathbf{p}}(\mathbf{C}\mathbf{R}_{\mathbf{hh}}\mathbf{C}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2(\mathbf{X}^{\mathbf{p}\mathrm{H}}\mathbf{X}^{\mathbf{p}})^{-1})\mathbf{X}^{\mathbf{p}\mathrm{H}}]^{-1} \\ &= \sigma_{\mathrm{p}}^{-2}\mathbf{R}_{\mathbf{hh}}\mathbf{C}^{\mathrm{H}}(\mathbf{C}\mathbf{R}_{\mathbf{hh}}\mathbf{C}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{I})^{-1}\mathbf{X}^{\mathbf{p}\mathrm{H}} \quad ,\end{aligned} \tag{C.27}$$

under the assumption that the pilot symbols have equal power, i.e. $\mathbf{X}^{\mathbf{p}\mathrm{H}}\mathbf{X}^{\mathbf{p}} = \sigma_{\mathrm{p}}^2\mathbf{I}$. Hence the linear MMSE estimate of the CFR vector is

$$\hat{\mathbf{h}}^{\mathrm{mmse}} = \mathbf{Q}\mathbf{y}^{\mathbf{p}} = \sigma_{\mathrm{p}}^{-2}\mathbf{R}_{\mathbf{hh}}\mathbf{C}^{\mathrm{H}}(\mathbf{C}\mathbf{R}_{\mathbf{hh}}\mathbf{C}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{I})^{-1}\mathbf{X}^{\mathbf{p}\mathrm{H}}\mathbf{y}^{\mathbf{p}}. \tag{C.28}$$

Under the finite CIR assumption, when $\mathbf{R}_{\mathbf{hh}}$ is decomposed according to (3.20),

$$\begin{aligned}\mathbf{Q} &= \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}(\mathbf{C}\mathbf{F}\mathbf{B}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}} + SNR_{\mathrm{p}}^{-1}\mathbf{I})^{-1}\mathbf{X}^{\mathbf{p}\mathrm{H}} \\ &= \sigma_{\mathrm{p}}^{-2}SNR_{\mathrm{p}}\mathbf{C}^{+}\mathbf{C}\mathbf{F}\mathbf{B}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}(SNR_{\mathrm{p}}\mathbf{C}\mathbf{F}\mathbf{B}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}} + \mathbf{I})^{-1}\mathbf{X}^{\mathbf{p}\mathrm{H}} \quad .\end{aligned} \tag{C.29}$$

Applying the matrix inversion identities [102],

$$\mathbf{A}(\mathbf{A}+\mathbf{I})^{-1} = \mathbf{I} - (\mathbf{A}+\mathbf{I})^{-1}, \tag{C.30}$$

$$\mathbf{I} - (\mathbf{U}\mathbf{V}\mathbf{U}^{\mathrm{H}} + \mathbf{I})^{-1} = \mathbf{U}(\mathbf{V}^{-1} + \mathbf{U}^{\mathrm{H}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{H}} = \mathbf{U}[\mathbf{I} - (\mathbf{U}^{\mathrm{H}}\mathbf{U})^{-1}(\mathbf{V} + (\mathbf{U}^{\mathrm{H}}\mathbf{U})^{-1})^{-1}](\mathbf{U}^{\mathrm{H}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{H}}, \tag{C.31}$$

we obtain

$$\begin{aligned}\mathbf{Q} &= \sigma_{\mathrm{p}}^{-2}\mathbf{C}^{+}[\mathbf{I} - (SNR_{\mathrm{p}}\mathbf{C}\mathbf{F}\mathbf{B}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}} + \mathbf{I})^{-1}]\mathbf{X}^{\mathbf{p}\mathrm{H}} \\ &= \sigma_{\mathrm{p}}^{-2}\mathbf{C}^{+}\mathbf{C}\mathbf{F}\mathbf{B}[\mathbf{I} - \mathbf{S}(SNR_{\mathrm{p}}\mathbf{R}_{\underline{\mathbf{hh}}} + \mathbf{S})^{-1}]\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\mathrm{H}} \\ &= \sigma_{\mathrm{p}}^{-2}\mathbf{F}\mathbf{B}[\mathbf{I} - SNR_{\mathrm{p}}^{-1}\mathbf{S}(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\mathrm{p}}^{-1}\mathbf{S})^{-1}]\mathbf{S}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\mathrm{H}} \quad ,\end{aligned} \tag{C.32}$$

where $\mathbf{S} = (\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\mathbf{F}\mathbf{B})^{-1}$.

C.4: Error correlation for CMMSE-type estimator

Let the CIR estimate be denoted according to (3.31) as

$$\hat{\underline{\mathbf{h}}} = \mathbf{G}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathbf{p}\mathrm{H}}\mathbf{y}^{\mathbf{p}}, \tag{C.33}$$

where $\mathbf{G} = \sigma_{\mathrm{p}}^{-2}(\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1})\mathbf{S}$, $\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}} = \overline{\mathbf{R}}_{\underline{hh}} + \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}$, and $\mathbf{y}^{\mathrm{p}} = \mathbf{X}^{\mathrm{p}}\mathbf{C}\mathbf{F}\mathbf{B}\underline{\mathbf{h}} + \mathbf{w}^{\mathrm{p}}$. Then

$$
\begin{aligned}
\mathbf{R}_{\underline{\varepsilon}\underline{\varepsilon}} &= \mathrm{E}[\underline{\boldsymbol{\varepsilon}}\underline{\boldsymbol{\varepsilon}}^{\mathrm{H}}] \\
&= \mathrm{E}[(\hat{\underline{\mathbf{h}}} - \underline{\mathbf{h}})(\hat{\underline{\mathbf{h}}} - \underline{\mathbf{h}})^{\mathrm{H}}] \\
&= \mathrm{E}(\hat{\underline{\mathbf{h}}}\hat{\underline{\mathbf{h}}}^{\mathrm{H}}) + \mathrm{E}(\underline{\mathbf{h}}\underline{\mathbf{h}}^{\mathrm{H}}) - \mathrm{E}(\hat{\underline{\mathbf{h}}}\underline{\mathbf{h}}^{\mathrm{H}}) - \mathrm{E}(\underline{\mathbf{h}}\hat{\underline{\mathbf{h}}}^{\mathrm{H}}) \\
&= \sigma_{\mathrm{p}}^4\mathbf{G}\mathbf{S}^{-1}\mathrm{E}(\underline{\mathbf{h}}\underline{\mathbf{h}}^{\mathrm{H}})\mathbf{S}^{-1}\mathbf{G}^{\mathrm{H}} + \mathbf{G}\mathbf{B}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{X}^{\mathrm{p}\,\mathrm{H}}\mathrm{E}(\mathbf{w}^{\mathrm{p}}\mathbf{w}^{\mathrm{p}\,\mathrm{H}})\mathbf{X}^{\mathrm{p}}\mathbf{C}\mathbf{F}\mathbf{B}\mathbf{G}^{\mathrm{H}} + \mathrm{E}(\underline{\mathbf{h}}\underline{\mathbf{h}}^{\mathrm{H}}) \\
&\quad - \sigma_{\mathrm{p}}^2\mathbf{G}\mathbf{S}^{-1}\mathrm{E}(\underline{\mathbf{h}}\underline{\mathbf{h}}^{\mathrm{H}}) - \sigma_{\mathrm{p}}^2\mathrm{E}(\underline{\mathbf{h}}\underline{\mathbf{h}}^{\mathrm{H}})\mathbf{S}^{-1}\mathbf{G}^{\mathrm{H}} \\
&= \sigma_{\mathrm{p}}^4\mathbf{G}\mathbf{S}^{-1}\mathbf{R}_{\underline{hh}}\mathbf{S}^{-1}\mathbf{G}^{\mathrm{H}} + \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^2\mathbf{G}\mathbf{S}^{-1}\mathbf{G}^{\mathrm{H}} + \mathbf{R}_{\underline{hh}} - \sigma_{\mathrm{p}}^2\mathbf{G}\mathbf{S}^{-1}\mathbf{R}_{\underline{hh}} - \sigma_{\mathrm{p}}^2\mathbf{R}_{\underline{hh}}\mathbf{S}^{-1}\mathbf{G}^{\mathrm{H}} \\
&= (\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1})\mathbf{R}_{\underline{hh}}(\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{S}) + \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}(\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1})\mathbf{S}(\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{S}) \\
&\quad + \mathbf{R}_{\underline{hh}} - (\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1})\mathbf{R}_{\underline{hh}} - \mathbf{R}_{\underline{hh}}(\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{S}) \\
&= -\bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{R}_{\underline{hh}}(\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{S}) + \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}(\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1})\mathbf{S}(\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{S}) \\
&\quad + \mathbf{R}_{\underline{hh}} - (\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1})\mathbf{R}_{\underline{hh}} \\
&= \bar{\sigma}_{\mathrm{w}}^4\sigma_{\mathrm{p}}^{-4}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{R}_{\underline{hh}}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{S} + \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}(\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1})\mathbf{S}(\mathbf{I} - \bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{S}) \\
&= \bar{\sigma}_{\mathrm{w}}^4\sigma_{\mathrm{p}}^{-4}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{R}_{\underline{\tilde{h}}\underline{\tilde{h}}}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{S} + \sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S} - 2\bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-4}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{S} \\
&= \mathbf{S}(\sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-2}\mathbf{S}^{-1} + \bar{\sigma}_{\mathrm{w}}^4\sigma_{\mathrm{p}}^{-4}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1}\mathbf{R}_{\underline{\tilde{h}}\underline{\tilde{h}}}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1} - 2\bar{\sigma}_{\mathrm{w}}^2\sigma_{\mathrm{w}}^2\sigma_{\mathrm{p}}^{-4}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1})\mathbf{S} \\
&= \mathbf{R}_{\underline{\tilde{\varepsilon}}\underline{\tilde{\varepsilon}}} - \mathbf{R}_{\underline{\eta}\underline{\eta}}
\end{aligned}
\quad , \qquad \text{(C.34)}
$$

where $\mathbf{R}_{\underline{\tilde{\varepsilon}}\underline{\tilde{\varepsilon}}} = SNR_{\mathrm{p}}^{-1}\mathbf{S}$ is correlation matrix of the CIR estimation error corresponding to the CLS estimator, $\mathbf{R}_{\underline{\eta}\underline{\eta}} = S\bar{N}R_{\mathrm{p}}^{-1}\mathbf{S}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1/2}(2\,SNR_{\mathrm{p}}^{-1}\mathbf{I} - S\bar{N}R_{\mathrm{p}}^{-1}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1/2}\mathbf{R}_{\underline{\tilde{h}}\underline{\tilde{h}}}\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1/2})\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}}^{-1/2}\mathbf{S}$ is the covariance reduction term for $\mathbf{R}_{\underline{\varepsilon}\underline{\varepsilon}}$, and $S\bar{N}R_{\mathrm{p}} = \bar{\sigma}_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^2$.

C.5: Proof of optimal pilot placement

To determine the optimal arrangement of pilot subcarriers (namely, matrix $\mathbf{C}^{\mathrm{H}}\mathbf{C}$ constituting $\mathbf{S}$) that minimises MSE (3.34), assume there exists a positive definite Hermitean matrix $\mathbf{R}$, such that $\mathbf{R}_{\underline{\eta}\underline{\eta}}$ in (3.33) becomes

$$
\mathbf{R}_{\underline{\eta}\underline{\eta}} = SNR_{\mathrm{p}}^{-2}\mathbf{S}(\mathbf{R} + SNR_{\mathrm{p}}^{-1}\mathbf{S})^{-1}\mathbf{S} = SNR_{\mathrm{p}}^{-2}\mathbf{S}\mathbf{U}(\mathbf{\Lambda} + SNR_{\mathrm{p}}^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{S}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{S}, \qquad \text{(C.35)}
$$

where $\mathbf{\Lambda}$ is a diagonal matrix with the eigenvalues of $\mathbf{R}$, and columns of $\mathbf{U}$ are the corresponding eigenvectors. In fact, this assumption is validated by the optimal estimator design when $\overline{\mathbf{R}}_{\underline{\tilde{h}}\underline{\tilde{h}}} = \mathbf{R}_{\underline{\tilde{h}}\underline{\tilde{h}}} = \mathbf{R} + SNR_{\mathrm{p}}^{-1}\mathbf{S}$ and $S\bar{N}R_{\mathrm{p}} = SNR_{\mathrm{p}}$. Then one can write according to the matrix inversion lemma [102] that

$$\begin{aligned}\mathbf{R}_{\underline{\underline{\varepsilon\varepsilon}}} &= \mathbf{R}_{\underline{\underline{\tilde{\varepsilon}\tilde{\varepsilon}}}} - \mathbf{R}_{\underline{\underline{\eta\eta}}} \\ &= \mathbf{U}[SNR_{\mathrm{p}}^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{S}\mathbf{U} - SNR_{\mathrm{p}}^{-2}\mathbf{U}^{\mathrm{H}}\mathbf{S}\mathbf{U}(\mathbf{\Lambda} + SNR_{\mathrm{p}}^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{S}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{S}\mathbf{U}]\mathbf{U}^{\mathrm{H}} \\ &= \mathbf{U}(\mathbf{\Lambda}^{-1} + SNR_{\mathrm{p}}\mathbf{U}^{\mathrm{H}}\mathbf{S}^{-1}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{H}}\end{aligned}, \tag{C.36}$$

and

$$\sigma_{\varepsilon}^2 = \mathrm{tr}(\mathbf{R}_{\underline{\underline{\varepsilon\varepsilon}}}) = \mathrm{tr}[(\mathbf{\Lambda}^{-1} + SNR_{\mathrm{p}}\mathbf{U}^{\mathrm{H}}\mathbf{S}^{-1}\mathbf{U})^{-1}]. \tag{C.37}$$

Using the trace version of the Cauchy-Schwarz inequality [131], one can show that

$$\sigma_{\varepsilon}^2\mathrm{tr}[\mathbf{\Lambda}^{-1} + SNR_{\mathrm{p}}\mathbf{U}^{\mathrm{H}}\mathbf{S}^{-1}\mathbf{U}] \geq (\mathrm{tr}[(\mathbf{\Lambda}^{-1} + SNR_{\mathrm{p}}\mathbf{U}^{\mathrm{H}}\mathbf{S}^{-1}\mathbf{U})^{-1/2}(\mathbf{\Lambda}^{-1} + SNR_{\mathrm{p}}\mathbf{U}^{\mathrm{H}}\mathbf{S}^{-1}\mathbf{U})^{1/2}])^2, \tag{C.38}$$

i.e.

$$\sigma_{\varepsilon}^2 \geq \frac{L^2}{\mathrm{tr}(\mathbf{\Lambda}^{-1}) + SNR_{\mathrm{p}}\mathrm{tr}(\mathbf{S}^{-1})}. \tag{C.39}$$

As $\mathbf{C}^{\mathrm{H}}\mathbf{C}$ is a singular diagonal matrix with P ones and $N-P$ zeros on the main diagonal and $\left\|((\mathbf{F}\mathbf{B})^{\mathrm{H}})_n\right\|^2 = L$ for $\forall n \in [0, N-1]$, then according to the definition of the diagonal matrix's trace [102]

$$\mathrm{tr}(\mathbf{S}^{-1}) = \mathrm{tr}[(\mathbf{F}\mathbf{B})^{\mathrm{H}}\mathbf{C}^{\mathrm{H}}\mathbf{C}\mathbf{F}\mathbf{B}] = \sum_{n=0}^{N-1}[\mathbf{C}^{\mathrm{H}}\mathbf{C}]_{k,k}\left\|((\mathbf{F}\mathbf{B})^{\mathrm{H}})_n\right\|^2 = P\,L. \tag{C.40}$$

Hence

$$\sigma_{\varepsilon}^2 \geq \frac{L^2}{\mathrm{tr}(\mathbf{\Lambda}^{-1}) + SNR_{\mathrm{p}}P\,L}, \tag{C.41}$$

where the equality holds if and only if $(\mathbf{\Lambda}^{-1} + SNR_{\mathrm{p}}\mathbf{U}^{\mathrm{H}}\mathbf{S}^{-1}\mathbf{U})^{-1}$ is a multiple of $\mathbf{\Lambda}^{-1} + SNR_{\mathrm{p}}\mathbf{U}^{\mathrm{H}}\mathbf{S}^{-1}\mathbf{U}$. This is possible if and only if $\mathbf{\Lambda}^{-1} + SNR_{\mathrm{p}}\mathbf{U}^{\mathrm{H}}\mathbf{S}^{-1}\mathbf{U}$ is a scaled identity matrix, i.e. when $\mathbf{\Lambda}^{-1} + SNR_{\mathrm{p}}\mathbf{U}^{\mathrm{H}}\mathbf{S}^{-1}\mathbf{U} = \beta\mathbf{I}$ and consequently $\mathbf{S}^{-1} = SNR_{\mathrm{p}}^{-1}\mathbf{U}(\beta\mathbf{I} - \mathbf{\Lambda}^{-1})\mathbf{U}^{\mathrm{H}}$, where the scalar multiplier β is selected to ensure positive semi-definiteness of $\beta\mathbf{I} - \mathbf{\Lambda}^{-1}$. Hence the optimal pilot arrangement in general is dependent on $\mathbf{R}$, which is channel-dependent in turn. However, as the eigenvectors of $\mathbf{S}$ for any $\mathbf{C}^{\mathrm{H}}\mathbf{C}$ combination in the absolute majority of cases differ from the eigenvectors of $\mathbf{R}$, exact optimal assignment of $\mathbf{S}^{-1}$ is impossible. Taking into account that $\sigma_{\varepsilon}^2 = \sigma_{\tilde{\varepsilon}}^2 - \sigma_{\eta}^2$, where $\sigma_{\tilde{\varepsilon}}^2$ has a bigger contribution than σ_{η}^2, one can select $\mathbf{S}^{-1}$ that minimises $\sigma_{\tilde{\varepsilon}}^2 = \mathrm{tr}(\mathbf{R}_{\underline{\underline{\tilde{\varepsilon}\tilde{\varepsilon}}}}) = SNR_{\mathrm{p}}^{-1}\mathrm{tr}(\mathbf{S})$ (3.13) and thus is optimal for the CLS estimator.

Applying repeatedly the trace version of the Cauchy-Schwarz inequality [131], it can be shown that

$$\sigma_{\tilde{\varepsilon}}^2\mathrm{tr}(\mathbf{S}^{-1}) \geq SNR_{\mathrm{p}}^{-1}[\mathrm{tr}(\mathbf{S}^{1/2}\mathbf{S}^{-1/2})]^2, \tag{C.42}$$

from where it follows that

$$\sigma_{\tilde{\varepsilon}}^2 \geq SNR_{\mathrm{p}}^{-1}P^{-1}L, \tag{C.43}$$

where the equality holds if and only if $\mathbf{S}$ is a scaled identity matrix, i.e. when the pilots are equispaced.

It is important to note that in a special case of both $\mathbf{S}$ and $\mathbf{R}$ being scaled identity matrices, $\mathbf{\Lambda}^{-1} + SNR_{\mathrm{p}}\mathbf{U}^{\mathrm{H}}\mathbf{S}^{-1}\mathbf{U}$ is straightforwardly recognised as the scaled identity matrix too. Thus, the CMMSE estimator has an optimal configuration, relying on the equispaced pilot pattern, from the standpoint of inequality (C.41).

C.6: Derivation of upper and lower bounds on MMSE of CMMSE estimator

Using the identity $\mathbf{A}(\mathbf{A}+\mathbf{I})^{-1} = \mathbf{I} - (\mathbf{A}+\mathbf{I})^{-1}$ [102], expression (3.40) can be rewritten as

$$\begin{aligned}\sigma_{\varepsilon}^2 &= \mathrm{tr}[\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2}(SNR_{\mathrm{p}}\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2} + \mathbf{I})^{-1}\mathbf{S}] \\ &= SNR_{\mathrm{p}}^{-1}\,\mathrm{tr}(\mathbf{S}) - SNR_{\mathrm{p}}^{-1}\mathrm{tr}[\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}(P\,SNR_{\mathrm{p}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1}\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}\mathbf{S}^{1/2}] \\ &= SNR_{\mathrm{p}}^{-1}(s^{-1} - u)\end{aligned}, \tag{C.44}$$

where $s = 1/\mathrm{tr}(\mathbf{S})$, and we let $u = \mathrm{tr}[\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}(P\,SNR_{\mathrm{p}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1}\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}\mathbf{S}^{1/2}]$.

From (C.44), it is evident that the maximum of MSE σ_{ε}^2 is attained when u is minimised. Applying the trace version of the Cauchy-Schwarz inequality [131] to the positive definite matrix $\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}(P\,SNR_{\mathrm{p}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1/2}[\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}(P\,SNR_{\mathrm{p}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1/2}]^{\mathrm{H}}$ in the trace argument of u, we obtain

$$\begin{aligned}u\,\mathrm{tr}[&\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}(P\,SNR_{\mathrm{p}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{1/2}[\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}(P\,SNR_{\mathrm{p}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{1/2}]^{\mathrm{H}}] \\ &\geq (\mathrm{tr}[\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}(P\,SNR_{\mathrm{p}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1/2}[\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}(P\,SNR_{\mathrm{p}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{1/2}]^{\mathrm{H}}])^2\end{aligned} \tag{C.45}$$

that can be reduced to

$$u \geq \frac{(\mathrm{tr}[\mathbf{S}])^2}{P\,SNR_{\mathrm{p}}\mathrm{tr}[\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}\mathbf{S}^{1/2}] + \mathrm{tr}[\mathbf{S}]}. \tag{C.46}$$

Taking into account that $\mathbf{U}_{\underline{\mathbf{hh}}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}} = P^{-1}\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2}$, $\mathrm{tr}(\mathbf{R}_{\underline{\mathbf{hh}}}) = 1$ and $s = 1/\mathrm{tr}(\mathbf{S})$, inequality (C.46) becomes

$$u \geq \frac{s^{-1}}{1 + s\,SNR_{\mathrm{p}}}, \tag{C.47}$$

where the equality holds if and only if $\mathbf{\Lambda}_{\underline{\mathbf{hh}}}$ is a scaled identity matrix. This requirement is in turn satisfied if and only if $\mathbf{R}_{\underline{\mathbf{hh}}} = (\mathrm{tr}[\mathbf{S}])^{-1}\mathbf{S} = s\mathbf{S}$. Hence the corresponding eigenvalue set is $\mathbf{\Lambda}_{\underline{\mathbf{hh}}} = s\,P^{-1}\mathbf{I}$.

Substitution of (C.47) into (C.44) yields

$$\sigma_{\varepsilon}^2 \leq s^{-1}SNR_{\mathrm{p}}^{-1} - \frac{s^{-1}SNR_{\mathrm{p}}^{-1}}{1 + s\,SNR_{\mathrm{p}}} = \frac{1}{1 + s\,SNR_{\mathrm{p}}}, \tag{C.48}$$

where the right hand side, $\hat{\sigma}_{\varepsilon}^2 = (1 + s\,SNR_{\mathrm{p}})^{-1}$, represents the upper bound on MMSE, such that for any λ_l distribution under the given SNR_{p} $\sigma_{\varepsilon}^2 \leq \hat{\sigma}_{\varepsilon}^2$.

The lower bound on $\sigma_\varepsilon^2 = P\,\mathrm{tr}[\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(P\,SNR_\mathrm{p}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\mathbf{I})^{-1}\mathbf{U}_{\underline{\mathbf{hh}}}^\mathrm{H}\mathbf{S}^{1/2}]$ (C.44) can also be determined using the Cauchy-Schwarz trace inequality [131]:

$$\sigma_\varepsilon^2\,\mathrm{tr}[\mathbf{S}^{-1/2}\mathbf{U}_{\underline{\mathbf{hh}}}(P\,SNR_\mathrm{p}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\mathbf{I})\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}^{-1}\mathbf{U}_{\underline{\mathbf{hh}}}^\mathrm{H}\mathbf{S}^{-1/2}]\geq P\,L^2\,, \tag{C.49}$$

$$\sigma_\varepsilon^2 \geq \frac{P\,L^2}{P\,SNR_\mathrm{p}\,\mathrm{tr}(\mathbf{S}^{-1})+\mathrm{tr}(\mathbf{S}^{-1/2}\mathbf{U}_{\underline{\mathbf{hh}}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}^{-1}\mathbf{U}_{\underline{\mathbf{hh}}}^\mathrm{H}\mathbf{S}^{-1/2})}\,, \tag{C.50}$$

$$\sigma_\varepsilon^2 \geq \frac{L^2}{P\,L\,SNR_\mathrm{p}+\mathrm{tr}(\mathbf{R}_{\underline{\mathbf{hh}}}^{-1})}\,, \tag{C.51}$$

where $\mathbf{R}_{\underline{\mathbf{hh}}}$ is assumed to be invertible.

The right part of the inequality (C.51) approaches zero when $\mathbf{R}_{\underline{\mathbf{hh}}}$ is close to singular. This is stronger manifested when $L-1$ eigenvalues of $\mathbf{R}_{\underline{\mathbf{hh}}}$ tend to zero. Hence one can conclude that the lower MMSE bound, $\breve{\sigma}_\varepsilon^2 \leq \sigma_\varepsilon^2$, is attained when $\mathbf{R}_{\underline{\mathbf{hh}}}$ and therefore $\mathbf{\Lambda}_{\underline{\mathbf{hh}}}$ are of unity rank.

C.7: Impact of mismatched noise variance design on MMSE

According to (3.34), a suboptimal CMMSE design is possible only for $\sigma_\eta^2 = \mathrm{tr}[\mathbf{R}_{\underline{\boldsymbol{\eta\eta}}}]\geq 0$. The sufficient condition to satisfy $\sigma_\eta^2 \geq 0$ is the positive semi-definiteness of $\mathbf{R}_{\underline{\boldsymbol{\eta\eta}}}$ in (3.33). To show that, one can write down:

$$\begin{aligned}
\mathbf{R}_{\underline{\boldsymbol{\eta\eta}}} &= \bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{S}\,\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}}^{-1/2}[2\sigma_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I}-\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}}^{-1/2}\mathbf{R}_{\underline{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}}\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}}^{-1/2}]\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}}^{-1/2}\mathbf{S}\\
&= \bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{S}[2\sigma_\mathrm{w}^2\sigma_\mathrm{p}^{-2}(\mathbf{R}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{S})^{-1}-\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}(\mathbf{R}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{S})^{-1}(\mathbf{R}_{\underline{\mathbf{hh}}}+\sigma_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{S})(\mathbf{R}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{S})^{-1}]\mathbf{S}\\
&= \bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{S}^{1/2}[2\sigma_\mathrm{w}^2\sigma_\mathrm{p}^{-2}(\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1}\\
&\quad-\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}(\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1}(\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2}+\sigma_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})(\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1}]\mathbf{S}^{1/2}
\end{aligned}\,. \tag{C.52}$$

Substitution of EVD $\mathbf{S}^{-1/2}\mathbf{R}_{\underline{\mathbf{hh}}}\mathbf{S}^{-1/2} = P\,\mathbf{U}_{\underline{\mathbf{hh}}}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}\mathbf{U}_{\underline{\mathbf{hh}}}^\mathrm{H}$ into (C.52) yields

$$\begin{aligned}
\mathbf{R}_{\underline{\boldsymbol{\eta\eta}}} &= \bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}[2\sigma_\mathrm{w}^2\sigma_\mathrm{p}^{-2}(P\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1}\\
&\quad-\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}(P\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1}(P\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\sigma_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})(P\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1}]\mathbf{U}_{\underline{\mathbf{hh}}}^\mathrm{H}\mathbf{S}^{1/2}\\
&= \bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}(P\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1/2}\mathbf{V}(P\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1/2}\mathbf{U}_{\underline{\mathbf{hh}}}^\mathrm{H}\mathbf{S}^{1/2}
\end{aligned}\,, \tag{C.53}$$

where

$$\begin{aligned}
\mathbf{V} &= 2\sigma_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I}-\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}(P\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1/2}(P\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I}+(\sigma_\mathrm{w}^2-\bar{\sigma}_\mathrm{w}^2)\sigma_\mathrm{p}^{-2}\mathbf{I})(P\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1/2}\\
&= 2\sigma_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I}-\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}[\mathbf{I}+(\sigma_\mathrm{w}^2-\bar{\sigma}_\mathrm{w}^2)\sigma_\mathrm{p}^{-2}(P\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1}]\\
&= (2\sigma_\mathrm{w}^2-\bar{\sigma}_\mathrm{w}^2)\sigma_\mathrm{p}^{-2}\mathbf{I}-(\sigma_\mathrm{w}^2-\bar{\sigma}_\mathrm{w}^2)\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-4}(P\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\bar{\sigma}_\mathrm{w}^2\sigma_\mathrm{p}^{-2}\mathbf{I})^{-1}\\
&= \sigma_\mathrm{p}^{-2}(\sigma_\mathrm{w}^2-\bar{\sigma}_\mathrm{w}^2)[(\sigma_\mathrm{w}^2-\bar{\sigma}_\mathrm{w}^2)^{-1}(2\sigma_\mathrm{w}^2-\bar{\sigma}_\mathrm{w}^2)\mathbf{I}-(P\,\bar{\sigma}_\mathrm{w}^{-2}\sigma_\mathrm{p}^2\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\mathbf{I})^{-1}]
\end{aligned}\,. \tag{C.54}$$

Using the identity $\mathbf{I}-(\mathbf{A}+\mathbf{I})^{-1} = \mathbf{A}\,(\mathbf{A}+\mathbf{I})^{-1}$ from [102], (C.54) becomes

$$\begin{aligned}\mathbf{V} &= \sigma_{\mathrm{p}}^{-2}(\sigma_{\mathrm{w}}^2 - \overline{\sigma}_{\mathrm{w}}^2)[(\sigma_{\mathrm{w}}^2 - \overline{\sigma}_{\mathrm{w}}^2)^{-1}\sigma_{\mathrm{w}}^2\mathbf{I} + \mathbf{I} - (P\overline{\sigma}_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^2\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1}] \\ &= \sigma_{\mathrm{p}}^{-2}(\sigma_{\mathrm{w}}^2 - \overline{\sigma}_{\mathrm{w}}^2)[(\sigma_{\mathrm{w}}^2 - \overline{\sigma}_{\mathrm{w}}^2)^{-1}\sigma_{\mathrm{w}}^2\mathbf{I} + P\overline{\sigma}_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^2\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(P\overline{\sigma}_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^2\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1}]\end{aligned}. \quad \text{(C.55)}$$

Note that $\mathbf{R}_{\underline{\boldsymbol{\eta\eta}}}$ (C.53) is positive semi-definite when the diagonal matrix $\mathbf{V}$ (C.55) is positive semi-definite. Consider two cases: $\overline{\sigma}_{\mathrm{w}}^2 < \sigma_{\mathrm{w}}^2$ and $\overline{\sigma}_{\mathrm{w}}^2 \geq \sigma_{\mathrm{w}}^2$. When $\overline{\sigma}_{\mathrm{w}}^2 < \sigma_{\mathrm{w}}^2$, the positive semi-definiteness of $\mathbf{V}$ is obvious. For $\overline{\sigma}_{\mathrm{w}}^2 \geq \sigma_{\mathrm{w}}^2$, one can show in general that positive semi-definiteness of $\mathbf{V}$ is satisfied if $(\overline{\sigma}_{\mathrm{w}}^2 - \sigma_{\mathrm{w}}^2)[(\overline{\sigma}_{\mathrm{w}}^2 - \sigma_{\mathrm{w}}^2)^{-1}\sigma_{\mathrm{w}}^2\mathbf{I} - P\overline{\sigma}_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^2\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(P\overline{\sigma}_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^2\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1}]$ is positive semi-definite. As the elements of $P\overline{\sigma}_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^2\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(P\overline{\sigma}_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^2\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1}$ are real and less than unity, this condition holds for $(\overline{\sigma}_{\mathrm{w}}^2 - \sigma_{\mathrm{w}}^2)^{-1}\sigma_{\mathrm{w}}^2 \geq 1$. Hence the defined suboptimal range of $\overline{\sigma}_{\mathrm{w}}^2$, such that $\sigma_{\eta}^2 \geq 0$ and $\sigma_{\varepsilon}^2 \leq \sigma_{\tilde{\varepsilon}}^2$, is $0 \leq \overline{\sigma}_{\mathrm{w}}^2 \leq 2\sigma_{\mathrm{w}}^2$. In all other cases of $\overline{\sigma}_{\mathrm{w}}^2 > 2\sigma_{\mathrm{w}}^2$, the definiteness of $\mathbf{R}_{\underline{\boldsymbol{\eta\eta}}}$ depends on $\mathbf{\Lambda}_{\underline{\mathbf{hh}}}$. It should be noted that in the ultimate case of $\overline{\sigma}_{\mathrm{w}}^2 >> 2\sigma_{\mathrm{w}}^2$, $(\overline{\sigma}_{\mathrm{w}}^2 - \sigma_{\mathrm{w}}^2)[(\overline{\sigma}_{\mathrm{w}}^2 - \sigma_{\mathrm{w}}^2)^{-1}\sigma_{\mathrm{w}}^2\mathbf{I} - P\overline{\sigma}_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^2\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(P\overline{\sigma}_{\mathrm{w}}^{-2}\sigma_{\mathrm{p}}^2\mathbf{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1}]$ becomes negative-definite with $\sigma_{\eta}^2 < 0$, and thus MSE experiences dramatic growth, proportional to the noise variance mismatch $\overline{\sigma}_{\mathrm{w}}^2 - \sigma_{\mathrm{w}}^2$.

C.8: Derivation of MSE of diagonal CMMSE estimator

Using the substitutions $\mathbf{S} = P^{-1}\mathbf{I}$ and $\overline{SNR}_{\mathrm{p}} = SNR_{\mathrm{p}}$ in (3.34), and taking into account that $\overline{\mathbf{R}}_{\underline{\mathbf{hh}}} = \mathrm{dg}(\mathbf{R}_{\underline{\mathbf{hh}}})$, one can express

$$\begin{aligned}\sigma_{\varepsilon,\mathrm{es}}^2 &= SNR_{\mathrm{p}}^{-1}P^{-1}L - SNR_{\mathrm{p}}^{-2}P^{-2}\mathrm{tr}(2\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}}^{-1} - \overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}}^{-1}\mathbf{R}_{\underline{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}}\overline{\mathbf{R}}_{\underline{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}}^{-1}) \\ &= P^{-1}L\,SNR_{\mathrm{p}}^{-1} + P^{-2}SNR_{\mathrm{p}}^{-2}\mathrm{tr}[(\overline{\mathbf{R}}_{\underline{\mathbf{hh}}} + P^{-1}SNR_{\mathrm{p}}^{-1}\mathbf{I})^{-2}(\mathbf{R}_{\underline{\mathbf{hh}}} + P^{-1}SNR_{\mathrm{p}}^{-1}\mathbf{I})] \\ &\quad - 2P^{-2}SNR_{\mathrm{p}}^{-2}\mathrm{tr}[(\overline{\mathbf{R}}_{\underline{\mathbf{hh}}} + P^{-1}SNR_{\mathrm{p}}^{-1}\mathbf{I})^{-1}] \\ &= P^{-1}L\,SNR_{\mathrm{p}}^{-1} + P^{-2}SNR_{\mathrm{p}}^{-2}\sum_{l=0}^{L-1}\frac{g_l + P^{-1}SNR_{\mathrm{p}}^{-1}}{(g_l + P^{-1}SNR_{\mathrm{p}}^{-1})^2} - 2P^{-2}SNR_{\mathrm{p}}^{-2}\sum_{l=0}^{L-1}\frac{1}{g_l + P^{-1}SNR_{\mathrm{p}}^{-1}} \\ &= \sum_{l=0}^{L-1}\left(P^{-1}SNR_{\mathrm{p}}^{-1} - \frac{P^{-2}SNR_{\mathrm{p}}^{-2}}{g_l + P^{-1}SNR_{\mathrm{p}}^{-1}}\right) \\ &= \sum_{l=0}^{L-1}\frac{P^{-1}SNR_{\mathrm{p}}^{-1}g_l}{g_l + P^{-1}SNR_{\mathrm{p}}^{-1}} \\ &= \sum_{l=0}^{L-1}\frac{g_l}{1 + P\,SNR_{\mathrm{p}}g_l}\end{aligned}. \quad \text{(C.56)}$$

C.9: Derivation of MSE of robust CMMSE estimator

Substitution of $\overline{\mathbf{R}}_{\underline{\mathbf{hh}}} = s\mathbf{S}$ and $\overline{SNR}_{\mathrm{p}} = SNR_{\mathrm{p}}$ into (3.34) yields

$$\begin{aligned}
\sigma_{\varepsilon}^2 &= \mathrm{tr}[SNR_{\mathrm{p}}^{-1}\mathbf{S} - 2\,SNR_{\mathrm{p}}^{-2}\mathbf{S}(s\mathbf{S} + SNR_{\mathrm{p}}^{-1}\mathbf{S})^{-1}\mathbf{S} + SNR_{\mathrm{p}}^{-2}\mathbf{S}(s\mathbf{S} + SNR_{\mathrm{p}}^{-1}\mathbf{S})^{-1}(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\mathrm{p}}^{-1}\mathbf{S})(s\mathbf{S} + SNR_{\mathrm{p}}^{-1}\mathbf{S})^{-1}\mathbf{S}] \\
&= \mathrm{tr}[SNR_{\mathrm{p}}^{-1}\mathbf{S} - 2\,SNR_{\mathrm{p}}^{-2}(s + SNR_{\mathrm{p}}^{-1})^{-1}\mathbf{S} + SNR_{\mathrm{p}}^{-2}(s + SNR_{\mathrm{p}}^{-1})^{-2}(\mathbf{R}_{\underline{\mathbf{hh}}} + SNR_{\mathrm{p}}^{-1}\mathbf{S})] \\
&= s^{-1}SNR_{\mathrm{p}}^{-1} - 2\,s^{-1}SNR_{\mathrm{p}}^{-2}(s + SNR_{\mathrm{p}}^{-1})^{-1} + SNR_{\mathrm{p}}^{-2}(s + SNR_{\mathrm{p}}^{-1})^{-2}\,\mathrm{tr}(\mathbf{R}_{\underline{\mathbf{hh}}}) + s^{-1}SNR_{\mathrm{p}}^{-3}(s + SNR_{\mathrm{p}}^{-1})^{-2} \\
&= \frac{SNR_{\mathrm{p}}^2(s^2 + 2\,s\,SNR_{\mathrm{p}}^{-1} + SNR_{\mathrm{p}}^{-2}) - 2\,SNR_{\mathrm{p}}(s + SNR_{\mathrm{p}}^{-1}) + s\,SNR_{\mathrm{p}}\mathrm{tr}(\mathbf{R}_{\underline{\mathbf{hh}}}) + 1}{s\,SNR_{\mathrm{p}}^3(s + SNR_{\mathrm{p}}^{-1})^2} \\
&= \frac{SNR_{\mathrm{p}}^2 s^2 + s\,SNR_{\mathrm{p}}\mathrm{tr}(\mathbf{R}_{\underline{\mathbf{hh}}})}{s\,SNR_{\mathrm{p}}^3(s + SNR_{\mathrm{p}}^{-1})^2} \\
&= \frac{s\,SNR_{\mathrm{p}}^2[s + SNR_{\mathrm{p}}^{-1} - SNR_{\mathrm{p}}^{-1} + SNR_{\mathrm{p}}^{-1}\mathrm{tr}(\mathbf{R}_{\underline{\mathbf{hh}}})]}{s\,SNR_{\mathrm{p}}^3(s + SNR_{\mathrm{p}}^{-1})^2} \\
&= \frac{1}{SNR_{\mathrm{p}}(s + SNR_{\mathrm{p}}^{-1})} + \frac{\mathrm{tr}(\mathbf{R}_{\underline{\mathbf{hh}}}) - 1}{SNR_{\mathrm{p}}^2(s + SNR_{\mathrm{p}}^{-1})^2} \\
&= \frac{1}{1 + s\,SNR_{\mathrm{p}}} + \frac{\mathrm{tr}(\mathbf{R}_{\underline{\mathbf{hh}}}) - 1}{(1 + s\,SNR_{\mathrm{p}})^2} \\
&= \frac{1}{1 + s\,SNR_{\mathrm{p}}} = \tilde{\sigma}_{\varepsilon}^2
\end{aligned} \quad . \text{(C.57)}$$

One can see that estimation MSE is indeed independent of the average power distribution within the actual CIR. It is interesting to note that $\sigma_{\varepsilon}^2 \le \tilde{\sigma}_{\varepsilon}^2$ only if the average CIR energy, $\mathrm{tr}(\mathbf{R}_{\underline{\mathbf{hh}}})$, does not exceed unity, given the design matrix $\overline{\mathbf{R}}_{\underline{\mathbf{hh}}} = s\mathbf{S}$. This circumstance is important to account for in case of imperfect received signal power gain at the input of the channel estimator. Under-amplification is recommended rather than the over-amplification.

C.10: Derivation of transfer function of infinite Wiener filter

Expanding (3.72) after substitution of (3.57) and (3.71) yields

$$\mathrm{E}\left[\sum_{k=0}^{\infty} f_{l,k}\,\tilde{\underline{h}}_{l,m-k}\,\tilde{\underline{h}}_{l,m-\nu}^{*}\right] - \mathrm{E}\left[\underline{h}_{l,m}\,\underline{h}_{l,m-\nu}^{*}\right] = 0 \tag{C.58}$$

that is reduced after swapping the sum and expectation operators to

$$\sum_{k=0}^{\infty} f_{l,k}\,\tilde{r}_{l,\nu-k} - r_{l,\nu} = 0\,, \tag{C.59}$$

where

$$r_{l,\nu} = \mathrm{E}[\underline{h}_{l,m}\,\underline{h}_{l,m-\nu}^{*}] = g_l\,r_{\nu} \tag{C.60}$$

is ACF of the lth CIR sample, and $\tilde{r}_{l,\nu} = \mathrm{E}[\underline{\tilde{h}}_{l,m}\underline{\tilde{h}}^{*}_{l,m-\nu}]$ is ACF of the input of the lth filter that can be expanded to

$$\begin{aligned}\tilde{r}_{l,\nu} &= \mathrm{E}[(\underline{h}_{l,m} + \underline{\tilde{\varepsilon}}_{l,m})(\underline{h}^{*}_{l,m-\nu} + \underline{\tilde{\varepsilon}}^{*}_{l,m-\nu})] \\ &= \mathrm{E}[\underline{h}_{l,m}\underline{h}^{*}_{l,m-\nu}] + \mathrm{E}[\underline{\tilde{\varepsilon}}_{l,m}\underline{\tilde{\varepsilon}}^{*}_{l,m-\nu}] \\ &= r_{l,\nu} + \rho_l\delta_\nu \end{aligned}, \tag{C.61}$$

where $\underline{h}_{l,m}$ and $\underline{\tilde{\varepsilon}}_{l,m}$ are mutually uncorrelated as stated in Subsection 3.4.1, $\delta_\nu = \begin{cases}1, & \nu = 0\\ 0, & \nu \neq 0\end{cases}$ is the Dirac delta function, and $\rho_l = \mathrm{E}[\underline{\tilde{\varepsilon}}_{l,m}\underline{\tilde{\varepsilon}}^{*}_{l,m}] = SNR_{\mathrm{p}}^{-1}[\mathbf{S}]_{l,l}$ (3.65) is the noise variance at the filter input.

Solution of $f_{l,k}$ utilises spectral factorisation. Applying the two-sided z-transform to the both sides of (C.59), we can put down that

$$\underbrace{\sum_{k=0}^{\infty} f_{l,k} z^{-k}}_{\Phi_l(z)} \underbrace{\sum_{\nu=-\infty}^{\infty} \tilde{r}_{l,\nu-k} z^{-(\nu-k)}}_{\tilde{S}_l(z)} - \underbrace{\sum_{\nu=-\infty}^{\infty} r_{l,\nu} z^{-\nu}}_{S_l(z)} = \Psi_l(z), \tag{C.62}$$

where $z = \mathrm{e}^{\mathrm{j}\omega}$, $\Phi_l(z)$ is the z-transform of the causal discrete-time sequence $f_{l,k}$, $\tilde{S}_l(z)$ and $S_l(z)$ are the two-sided z-transforms of ACFs $\tilde{r}_{l,\nu}$ and $r_{l,\nu}$, respectively, and $\Psi_l(z)$ is the two-sided z-transform of the strictly anticausal sequence, which may take non-zero values only in the interval $\nu \in (-\infty, 0)$. Here the term $\tilde{S}_l(z)$ can be factored [139] as

$$\tilde{S}_l(z) = \mu_l Q_l(z) Q_l^{*}(1/z^{*}), \tag{C.63}$$

where $Q_l(z) = \sum_{k=0}^{\infty} q_{l,k} z^{-k}$ is the z-transform of the causal sequence $q_{l,k}$, $k \in [0, \infty)$, $Q_l^{*}(1/z^{*}) = \sum_{k=0}^{\infty} q^{*}_{l,-k}(z^{*})^{k}$ is the (two-sided) z-transform of the anticausal sequence $q^{*}_{l,k'}$, $k' \in (-\infty, 0]$, and $\mu_l = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln \tilde{S}_l(\mathrm{e}^{\mathrm{j}\omega})\,\mathrm{d}\omega\right)$. By definition of the spectral factorisation [139], both $q_{l,k}$ and $q^{*}_{l,k'}$ are monic, i.e. $q_{l,0} = q^{*}_{l,0} = 1$, hence μ_l is interpreted as the gain factor.

As $\tilde{S}_l(z)$ is the two-sided z-transform of $\tilde{r}_{l,\nu}$ (C.61), it can be expressed through $S_l(z)$ in the form of

$$\tilde{S}_l(z) = S_l(z) + \rho_l. \tag{C.64}$$

Substitution of (C.63) and $S_l(z) = \tilde{S}_l(z) - \rho_l$ into (C.62) yields:

$$\mu_l \Phi_l(z) Q_l(z) Q_l^{*}(1/z^{*}) - \mu_l Q_l(z) Q_l^{*}(1/z^{*}) + \rho_l = \Psi_l(z). \tag{C.65}$$

We can identify the causal parts on both sides of (C.65) after taking the term $Q_l^{*}(1/z^{*})$ outside:

$$\left\{ Q_l^{*}(1/z^{*}) \left[\mu_l \Phi_l(z) Q_l(z) - \mu_l Q_l(z) + \frac{\rho_l}{Q_l^{*}(1/z^{*})} \right] \right\}_{+} = 0, \tag{C.66}$$

where $\{.\}_+$ indicates that the causal part has been taken of the inverse two-sided z-transform of the function in braces. The operator $\{.\}_+$ has the following properties [140] in the context of (C.65):

$$\{\Psi_l(z)\}_+ = 0, \tag{C.67}$$

$$\{\Phi_l(z)\}_+ = \Phi_l(z), \tag{C.68}$$

$$\{Q_l(z)\}_+ = Q_l(z), \tag{C.69}$$

$$\{\Phi_l(z)Q_l(z)\}_+ = \{\{\Phi_l(z)\}_+\{Q_l(z)\}_+\}_+ = \{\Phi_l(z)\}_+\{Q_l(z)\}_+ = \Phi_l(z)Q_l(z), \tag{C.70}$$

$$\{Q_l^*(1/z^*)\}_+ = 1, \tag{C.71}$$

$$\{1/Q_l^*(1/z^*)\}_+ = 1, \tag{C.72}$$

where the last two identities hold as $Q_l^*(1/z^*)$ corresponds to the monic non-strictly anticausal sequence $q_{l,k'}^*$, which is non-zero in the interval $k' \in (-\infty, 0]$.

Using properties (C.67)-(C.72), we can proceed to the solution of the equation (C.66) in the following sequence:

$$\left\{\mu_l\Phi_l(z)Q_l(z) - \mu_l Q_l(z) + \frac{\rho_l}{Q_l^*(1/z^*)}\right\}_+ = 0, \tag{C.73}$$

$$\mu_l\Phi_l(z)Q_l(z) - \mu_l Q_l(z) + \rho_l = 0 \tag{C.74}$$

that results in

$$\Phi_l(z) = 1 - \frac{\rho_l}{\mu_l Q_l(z)}, \tag{C.75}$$

where $\Phi_l(z)$ is the Wiener filter transfer function, which is optimal in the MMSE sense. Later on we will be using just the normalised angular frequency ω as the argument of $\Phi_l(e^{j\omega})$ and $\tilde{S}_l(e^{j\omega})$ evaluated on the unit circle.

C.11: Derivation of MSE of infinite Wiener filter

One can compute the variance of the error at the output of the infinite Wiener filter as

$$\begin{aligned}
\sigma_{\underline{\varepsilon}l}^2 &= \mathrm{E}[\underline{\varepsilon}_{l,m}\underline{\varepsilon}_{l,m}^*] = \mathrm{E}\left[\left(\sum_{k=0}^{\infty} f_{l,k}\underline{\tilde{h}}_{l,m-k} - \underline{h}_{l,m}\right)\left(\sum_{\nu=0}^{\infty} f_{l,\nu}^*\underline{\tilde{h}}_{l,m-\nu}^* - \underline{h}_{l,m}^*\right)\right] \\
&= \mathrm{E}\left[\left(\sum_{k=0}^{\infty} f_{l,k}\underline{\tilde{h}}_{l,m-k}\right)\left(\sum_{\nu=0}^{\infty} f_{l,\nu}^*\underline{\tilde{h}}_{l,m-\nu}^*\right)\right] - \mathrm{E}\left[\sum_{k=0}^{\infty} f_{l,k}\underline{\tilde{h}}_{l,m-k}\underline{h}_{l,m}^*\right] - \mathrm{E}\left[\sum_{\nu=0}^{\infty} f_{l,\nu}^*\underline{\tilde{h}}_{l,m-\nu}^*\underline{h}_{l,m}\right] + \mathrm{E}\left[\underline{h}_{l,m}\underline{h}_{l,m}^*\right] \\
&= \sum_{k=0}^{\infty}\sum_{\nu=0}^{\infty} f_{l,k}f_{l,\nu}^*\mathrm{E}\left[\underline{\tilde{h}}_{l,m-k}\underline{\tilde{h}}_{l,m-\nu}^*\right] - \sum_{k=0}^{\infty} f_{l,k}\mathrm{E}\left[\underline{h}_{l,m-k}\underline{h}_{l,m}^*\right] - \sum_{\nu=0}^{\infty} f_{l,\nu}^*\mathrm{E}\left[\underline{h}_{l,m-\nu}^*\underline{h}_{l,m}\right] + \mathrm{E}\left[\underline{h}_{l,m}\underline{h}_{l,m}^*\right] \\
&= g_l + \sum_{k=0}^{\infty}\sum_{\nu=0}^{\infty} f_{l,k}f_{l,\nu}^*\tilde{r}_{l,\nu-k} - \sum_{k=0}^{\infty} f_{l,k}r_{l,-k} - \sum_{\nu=0}^{\infty} f_{l,\nu}^* r_{l,\nu}
\end{aligned} \quad . \tag{C.76}$$

Expressing $r_{l,k}$ and $\tilde{r}_{l,k}$ as the inverse discrete-time Fourier transform of the corresponding Doppler PSDs, $S_l(\omega)$ and $\tilde{S}_l(\omega)$, and taking into account the symmetry property of ACF, $r_{l,-k} = r_{l,k}^*$, (C.76) becomes

$$\begin{aligned}\sigma_{\underline{\varepsilon}l}^2 &= g_l + \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{k=0}^{\infty}\sum_{\nu=0}^{\infty} f_{l,k} f_{l,\nu}^* \tilde{S}_l(\omega)\mathrm{e}^{\mathrm{j}\omega(\nu-k)}\mathrm{d}\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{k=0}^{\infty} f_{l,k} S_l(\omega)\mathrm{e}^{-\mathrm{j}\omega k}\mathrm{d}\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{\nu=0}^{\infty} f_{l,\nu}^* S_l(\omega)\mathrm{e}^{\mathrm{j}\omega\nu}\mathrm{d}\omega \\ &= g_l + \frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_l(\omega)\Phi_l^*(\omega)\tilde{S}_l(\omega)\mathrm{d}\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_l(\omega)S_l(\omega)\mathrm{d}\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi_l^*(\omega)S_l(\omega)\mathrm{d}\omega \end{aligned} \quad . \quad (C.77)$$

C.12: Robust design of infinite Wiener filter

Let $\overline{\Phi}_l(\omega)$ be the transfer function of the filter designed for the Doppler PSD $\overline{S}(\omega)$, so that in analogy with (3.77), the spectral factorisation can be applied on the unit circle:

$$\overline{\tilde{S}}_l(\omega) = \overline{\mu}_l\,\overline{Q}_l(\omega)\overline{Q}_l^*(\omega), \quad (C.78)$$

where according to (3.75) and (3.76)

$$\overline{\mu}_l = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\overline{\tilde{S}}_l(\omega)\mathrm{d}\omega\right) = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln[g_l\overline{S}(\omega)+\rho_l]\mathrm{d}\omega\right). \quad (C.79)$$

Then in the conditions of the design mismatch (non-optimal Doppler PSD setting $\overline{S}(\omega)$), i.e. $\overline{\Phi}_l(\omega) \neq \Phi_l(\omega)$, where $\Phi_l(\omega)$ is based on the true $S(\omega)$, MSE at the output of the *l*th filter is found according to (3.78), where $\Phi_l(\omega)$ is replaced by

$$\overline{\Phi}_l(\omega) = 1 - \frac{\rho_l}{\overline{\mu}_l\overline{Q}_l(\omega)}. \quad (C.80)$$

Substituting (C.80) and (3.75) into (3.78) and using (C.78) yields

$$\begin{aligned}\sigma_{\underline{\varepsilon}l}^2 &= g_l + \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(1-\frac{\rho_l}{\overline{\mu}_l\overline{Q}_l(\omega)}\right)\left(1-\frac{\rho_l}{\overline{\mu}_l\overline{Q}_l^*(\omega)}\right)\left(S_l(\omega)+\rho_l\right)\mathrm{d}\omega \\ &\quad - \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(1-\frac{\rho_l}{\overline{\mu}_l\overline{Q}_l(\omega)}\right)S_l(\omega)\mathrm{d}\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(1-\frac{\rho_l}{\overline{\mu}_l\overline{Q}_l^*(\omega)}\right)S_l(\omega)\mathrm{d}\omega \\ &= g_l + \frac{\rho_l}{2\pi}\int_{-\pi}^{\pi}\left(1-\frac{\rho_l}{\overline{\mu}_l\overline{Q}_l(\omega)}\right)\left(1-\frac{\tilde{S}_l(\omega)}{\overline{\mu}_l\overline{Q}_l^*(\omega)}\right)\mathrm{d}\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(1-\frac{\rho_l}{\overline{\mu}_l\overline{Q}_l^*(\omega)}\right)S_l(\omega)\mathrm{d}\omega \\ &= g_l + \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\rho_l - \frac{\rho_l^2}{\overline{\mu}_l\overline{Q}_l(\omega)} - \frac{\rho_l\tilde{S}_l(\omega)}{\overline{\mu}_l\overline{Q}_l^*(\omega)} + \frac{\rho_l^2\tilde{S}_l(\omega)}{\overline{\mu}_l\overline{\tilde{S}}_l(\omega)} - S_l(\omega) + \frac{\rho_l S_l(\omega)}{\overline{\mu}_l\overline{Q}_l^*(\omega)}\right)\mathrm{d}\omega \\ &= g_l + \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\rho_l - S_l(\omega) - \frac{\rho_l^2}{\overline{\mu}_l\overline{Q}_l(\omega)} - \frac{\rho_l^2}{\overline{\mu}_l\overline{Q}_l^*(\omega)} + \frac{\rho_l^2\tilde{S}_l(\omega)}{\overline{\mu}_l\overline{\tilde{S}}_l(\omega)}\right)\mathrm{d}\omega \\ &= \rho_l\left(1 - \frac{2\rho_l}{\overline{\mu}_l} + \frac{\rho_l}{2\pi\overline{\mu}_l}\int_{-\pi}^{\pi}\frac{\tilde{S}_l(\omega)}{\overline{\tilde{S}}_l(\omega)}\mathrm{d}\omega\right) \end{aligned}, \quad (C.81)$$

where

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\overline{Q}_l(\omega)\right]^{-1}\mathrm{d}\omega=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\overline{Q}_l^{*}(\omega)\right]^{-1}\mathrm{d}\omega=1 \tag{C.82}$$

due to the monicity property of $\overline{Q}_l(\omega)$, and

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{S}_l(\omega)\,\mathrm{d}\omega=\frac{g_l}{2\pi}\int_{-\pi}^{\pi}\mathrm{S}(\omega)\,\mathrm{d}\omega=g_l \tag{C.83}$$

as ACF r_v of the Doppler process with PSD $S(\omega)$ is monic too. Using the equalities (3.75) and (3.76) in (C.81), we obtain

$$\begin{aligned}\sigma_{\underline{\varepsilon}l}^2&=\rho_l\left(1-\frac{2\rho_l}{\overline{\mu}_l}+\frac{\rho_l}{2\pi\,\overline{\mu}_l}\int_{-\pi}^{\pi}\frac{g_lS(\omega)+\rho_l}{g_l\overline{S}(\omega)+\rho_l}\mathrm{d}\omega\right)\\&=\rho_l\left(1-\frac{\rho_l}{\overline{\mu}_l}+\frac{\rho_l}{2\pi\,\overline{\mu}_l}\int_{-\pi}^{\pi}\left(\frac{g_lS(\omega)+\rho_l}{g_l\overline{S}(\omega)+\rho_l}-1\right)\mathrm{d}\omega\right)\\&=\rho_l\left(1-\frac{\rho_l}{\overline{\mu}_l}+\frac{g_l\rho_l}{2\pi\,\overline{\mu}_l}\int_{-\pi}^{\pi}\frac{S(\omega)-\overline{S}(\omega)}{g_l\overline{S}(\omega)+\rho_l}\mathrm{d}\omega\right)\end{aligned}\quad. \tag{C.84}$$

If design of the filterbank is optimal, i.e. $\overline{S}(\omega)=S(\omega)$ and $\overline{\mu}_l=\mu_l$, then (C.84) is reduced to the MMSE expression

$$\sigma_{\underline{\varepsilon}l}^2=\rho_l\left(1-\frac{\rho_l}{\mu_l}\right), \tag{C.85}$$

where

$$\mu_l=\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln[g_lS(\omega)+\rho_l]\,\mathrm{d}\omega\right). \tag{C.86}$$

For an arbitrary Doppler PSD $S(\omega)\neq\overline{S}(\omega)$, bandlimited in the interval $\omega\in[-\tilde{\omega}_{\mathrm{D}},\tilde{\omega}_{\mathrm{D}})$, the last term in (C.84) is equal to zero if and only if

$$\overline{S}(\omega)=\begin{cases}\pi/\tilde{\omega}_{\mathrm{D}}, & |\omega|\le\tilde{\omega}_{\mathrm{D}}\\ 0, & \text{otherwise}\end{cases}. \tag{C.87}$$

Thus, the robust Wiener filter design is possible, assuming the Doppler spectrum flat and bandlimited by $\tilde{\omega}_{\mathrm{D}}$. Then the error variance at the output of the *l*th filter is the same for any Doppler spectrum, bandlimited by $\tilde{\omega}_{\mathrm{D}}$, and is equal to

$$\sigma_{\underline{\varepsilon}l}^2=\rho_l\left(1-\frac{\rho_l}{\overline{\mu}_l}\right), \tag{C.88}$$

where

$$\begin{aligned}\overline{\mu}_l &= \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln[g_l\overline{S}(\omega)+\rho_l]\,\mathrm{d}\omega\right)\\ &= \exp\left(\frac{1}{2\pi}\int_{-\tilde{\omega}_\mathrm{D}}^{\tilde{\omega}_\mathrm{D}}\ln\left[\frac{\pi g_l}{\tilde{\omega}_\mathrm{D}}+\rho_l\right]\mathrm{d}\omega+\frac{1}{\pi}\int_{\tilde{\omega}_\mathrm{D}}^{\pi}\ln\rho_l\,\mathrm{d}\omega\right)\\ &= \exp\left(\frac{\tilde{\omega}_\mathrm{D}}{\pi}\ln\left[\frac{\pi g_l}{\tilde{\omega}_\mathrm{D}}+\rho_l\right]+\left(1-\frac{\tilde{\omega}_\mathrm{D}}{\pi}\right)\ln\rho_l\right)\\ &= \exp\left(\frac{\tilde{\omega}_\mathrm{D}}{\pi}\ln\left[\frac{\pi g_l}{\tilde{\omega}_\mathrm{D}\rho_l}+1\right]+\ln\rho_l\right)\\ &= \rho_l\left(\frac{\pi g_l}{\tilde{\omega}_\mathrm{D}\rho_l}+1\right)^{\tilde{\omega}_\mathrm{D}/\pi}\end{aligned}\quad . \tag{C.89}$$

Substitution of (C.89) into (C.88) yields

$$\sigma_{\underline{\varepsilon}l}^2 = \rho_l\left[1-\left(g_l\rho_l^{-1}\Omega+1\right)^{-\Omega^{-1}}\right], \tag{C.90}$$

where $\Omega = \pi/\tilde{\omega}_\mathrm{D}$.

Note that the robust design's MSE (C.90) is also interpreted as MMSE (C.85) if the true Doppler spectrum of the filtered sequence is flat and bandlimited by $\tilde{\omega}_\mathrm{D}$, i.e. $S(\omega) = \overline{S}(\omega)$ (C.87).

C.13: Derivation of upper and lower bounds on MSE of robust CLS-MMSE estimator

It is obvious from the MSE expression (3.83),

$$\sigma_\varepsilon^2 = \sum_{l=0}^{L-1}\rho_l\left[1-\left(g_l\rho_l^{-1}\Omega+1\right)^{-\Omega^{-1}}\right], \tag{C.91}$$

that the maximum of σ_ε^2 corresponds to the minimum of

$$\psi = \sum_{l=0}^{L-1}c_l\left(\Omega\,SNR_\mathrm{p}\,\frac{g_l}{c_l}+1\right)^{-\Omega^{-1}}, \tag{C.92}$$

which in turn matches the minimum of

$$\theta = \sum_{l=0}^{L-1}c_l^{\Omega}\left(\Omega\,SNR_\mathrm{p}\,\frac{g_l}{c_l}+1\right)^{-1} \tag{C.93}$$

as the power, i.e. $(.)^{\Omega^{-1}}$ in our case, is a monotonically increasing function of the argument. Assuming that $g_l \neq 0$, $\forall l \in [0, L-1]$ for the high SNR regime (when $SNR_\mathrm{p} \to \infty$), (C.93) can be approximated as

$$\theta \approx \frac{1}{\Omega\, SNR_{\mathrm{p}}} \sum_{l=0}^{L-1} \frac{c_l^{\Omega+1}}{g_l} . \tag{C.94}$$

Then, applying the Cauchy's inequality [103], one can show that

$$\theta \sum_{l=0}^{L-1} \frac{g_l}{c_l^{\Omega+1}} \geq \left[\left(\frac{1}{\Omega\, SNR_{\mathrm{p}}} \right)^{1/2} \sum_{l=0}^{L-1} \left(\frac{c_l^{\Omega+1}}{g_l} \right)^{1/2} \left(\frac{g_l}{c_l^{\Omega+1}} \right)^{1/2} \right]^2 . \tag{C.95}$$

Hence

$$\theta \geq \frac{L^2}{\Omega\, SNR_{\mathrm{p}}} \left(\sum_{l=0}^{L-1} \frac{g_l}{c_l^{\Omega+1}} \right)^{-1} , \tag{C.96}$$

where the equality holds if and only if

$$g_l = \beta\, c_l^{\Omega+1} , \tag{C.97}$$

where $\beta = \left(\sum_{l=0}^{L-1} c_l^{\Omega+1} \right)^{-1}$.

Substitution of (C.97) into (C.94) and (C.92) yields expressions:

$$\theta_{\min} = \frac{L}{\Omega\, SNR_{\mathrm{p}}} \sum_{l=0}^{L-1} c_l^{\Omega+1} , \tag{C.98}$$

$$\psi_{\min} = \sum_{l=0}^{L-1} c_l \left(\Omega\, SNR_{\mathrm{p}} \beta\, c_l^{\Omega} + 1 \right)^{-\Omega^{-1}} , \tag{C.99}$$

where the minimum of (C.98) and hence (C.99) is attained if and only if the pilot arrangement is equispaced, i.e. when $c_l = P^{-1}$, $\forall l \in [0, L-1]$. The latter conclusion follows from the equivalence of the minimums of $\sum_{l=0}^{L-1} c_l^{\Omega+1}$ and $\sum_{l=0}^{L-1} c_l$, and the trace version of the Cauchy-Schwarz inequality [131]:

$$\mathrm{tr}(\mathbf{S}) \geq L^2 / \mathrm{tr}(\mathbf{S}^{-1}) , \tag{C.100}$$

which can be shown to have the reduced form of

$$\sum_{l=0}^{L-1} c_l \geq L/P , \tag{C.101}$$

where the equality holds if and only if $c_l = P^{-1}$. (Here the definition of the diagonal matrix's trace [102] has been used to obtain $\mathrm{tr}(\mathbf{S}^{-1}) = \mathrm{tr}[(\mathbf{FB})^{\mathrm{H}} \mathbf{C}^{\mathrm{H}} \mathbf{C} \mathbf{FB}] = PL$.)

Thus, for a general (not only equispaced) pilot arrangement, it holds true that

$$\theta \geq \frac{L^2}{\Omega\, SNR_{\mathrm{p}} P^{\Omega+1}} \tag{C.102}$$

and, taking into account $\beta = P^{\Omega+1}/L$ in (C.99),

$$\psi \geq \frac{L}{P}\left(\frac{\Omega P\, SNR_{\mathrm{p}}}{L}+1\right)^{-\Omega^{-1}}, \tag{C.103}$$

with the equality only under the condition (C.97), applied to the equispaced pilot pattern.

It should be noted that for the equispaced pilot positioning, equations (C.97), (C.102) and (C.103) can be derived with no need of assumptions and the approximation introduced in (C.94).

Using (C.97) in (C.91), we can find the upper bound on MSE:

$$\widehat{\sigma}_{\varepsilon}^{2} = SNR_{\mathrm{p}}^{-1}\sum_{l=0}^{L-1} c_l\left[1-(\Omega\, SNR_{\mathrm{p}}\beta\, c_l^{\Omega}+1)^{-\Omega^{-1}}\right] = SNR_{\mathrm{p}}^{-1}\left[\sum_{l=0}^{L-1} c_l - \psi_{\min}\right], \tag{C.104}$$

such that for any g_l distribution $\sigma_{\varepsilon}^{2} \leq \widehat{\sigma}_{\varepsilon}^{2}$, where $\sigma_{\varepsilon}^{2} = \widehat{\sigma}_{\varepsilon}^{2}$ corresponds to the worst-case scenario of the estimator performance.

It has been established by means of the series of equations (C.98)-(C.103) that the subtracted term in (C.104), $\psi_{\min}$, attains minimum when the pilot subcarriers are equispaced. Thus, after substitution of the right side of (C.103) instead of $\psi_{\min}$ in (C.104), one can show that

$$\widehat{\sigma}_{\varepsilon}^{2} \leq \frac{1}{SNR_{\mathrm{p}}}\left[\frac{1}{s}-\frac{L}{P}\left(\frac{\Omega P\, SNR_{\mathrm{p}}}{L}+1\right)^{-\Omega^{-1}}\right], \tag{C.105}$$

where $s = \left(\sum_{l=0}^{L-1} c_l\right)^{-1}$, and the equality holds if and only if the pilot pattern is equispaced. Taking this result into account and recalling inequality (C.101) provides an insight of the upper bound (C.105) being the smallest for the equispaced pilot scheme. In our analysis, we will use the approximated upper MSE bound given by the right side of (C.105) rather than the precise formula (C.104).

The lower bound on MSE (C.91), signifying the best performance of the estimator, can be assessed by searching for the utmost value of θ (C.93). One can see that θ (and hence ψ (C.92)) reaches maximum when $g_l = 0$ for all the PDP samples except one, i.e. when the channel is flat fading. An intuitive reasoning is that in such a case MSE is contributed by only one filter in the Wiener filterbank, whereas the other ones fully block their outputs by setting $g_l = 0$ in (3.67). Thus, the largest ψ can be expressed as

$$\psi_{\max} = c_{l_0}\left(\Omega\, SNR_{\mathrm{p}} c_{l_0}^{-1}+1\right)^{-\Omega^{-1}} + \sum_{\substack{l=0 \\ l\neq l_0}}^{L-1} c_l \approx \left(\Omega\, SNR_{\mathrm{p}}\right)^{-\Omega^{-1}} c_{l_0}^{\Omega^{-1}+1} + \sum_{\substack{l=0 \\ l\neq l_0}}^{L-1} c_l\,, \tag{C.106}$$

where $l_0 = \arg\min_{l\in[0,L-1]} c_l$ identifies the most precise filter in the bank, and the lower bound on MSE is found as

$$\breve{\sigma}_{\varepsilon}^{2} = SNR_{\mathrm{p}}^{-1}\left[\sum_{l=0}^{L-1} c_l - \psi_{\max}\right] = SNR_{\mathrm{p}}^{-1} c_{l_0}\left[1-\left(\Omega\, SNR_{\mathrm{p}} c_{l_0}^{-1}+1\right)^{-\Omega^{-1}}\right]. \tag{C.107}$$

Appendix D

D.1: Characterisation of detection error

In order to obtain an expression for the detection error on the *n*th subcarrier, we substitute equation (2.47) into (4.5), assuming $\mathbf{H}_{n,m}$ and $\mathbf{E}_{n,m}$ to be full rank (equal to N_{tx}), and use the Searle set of inversion identities [138]. The output of the linear ZF detector is then described as

$$\begin{aligned}
\hat{\mathbf{x}}_{n,m} &= \hat{\mathbf{H}}^{+}_{n,m}\mathbf{y}_{n,m} \\
&= \mathbf{x}_{n,m} + [(\mathbf{H}_{n,m} + \mathbf{E}_{n,m})^{+} - \mathbf{H}^{+}_{n,m}]\mathbf{H}_{n,m}\mathbf{x}_{n,m} + (\mathbf{H}_{n,m} + \mathbf{E}_{n,m})^{+}\mathbf{w}_{n,m} \\
&= \mathbf{x}_{n,m} - [(\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m})^{-1}\mathbf{H}^{\mathrm{H}}_{n,m} - ((\mathbf{H}_{n,m} + \mathbf{E}_{n,m})^{\mathrm{H}}(\mathbf{H}_{n,m} + \mathbf{E}_{n,m}))^{-1}(\mathbf{H}_{n,m} + \mathbf{E}_{n,m})^{\mathrm{H}}]\mathbf{H}_{n,m}\mathbf{x}_{n,m} \\
&\quad + [(\mathbf{H}_{n,m} + \mathbf{E}_{n,m})^{\mathrm{H}}(\mathbf{H}_{n,m} + \mathbf{E}_{n,m})]^{-1}(\mathbf{H}_{n,m} + \mathbf{E}_{n,m})^{\mathrm{H}}\mathbf{w}_{n,m} \\
&= \mathbf{x}_{n,m} - (\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m})^{-1}[\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m} - \mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m}((\mathbf{H}_{n,m} + \mathbf{E}_{n,m})^{\mathrm{H}}\mathbf{H}_{n,m} + (\mathbf{H}_{n,m} + \mathbf{E}_{n,m})^{\mathrm{H}}\mathbf{E}_{n,m})^{-1}(\mathbf{H}_{n,m} \\
&\quad + \mathbf{E}_{n,m})^{\mathrm{H}}\mathbf{H}_{n,m}]\mathbf{x}_{n,m} + [(\mathbf{H}_{n,m} + \mathbf{E}_{n,m})^{\mathrm{H}}\mathbf{H}_{n,m} + (\mathbf{H}_{n,m} + \mathbf{E}_{n,m})^{\mathrm{H}}\mathbf{E}_{n,m}]^{-1}(\mathbf{H}_{n,m} + \mathbf{E}_{n,m})^{\mathrm{H}}\mathbf{w}_{n,m}
\end{aligned} \quad . \qquad \text{(D.1)}$$

The resultant expression can be further simplified, assuming that the system is being optimised for operation in the high SNR region. The high SNR regime is often of practical interest, especially when performance analysis is done using the closed-form error probability functions [141]. The simplification of (D.1) relies on the approximation $\mathbf{H}_{n,m} + \mathbf{E}_{n,m} \approx \mathbf{H}_{n,m}$ that is valid when the channel estimation errors, forming $\mathbf{E}_{n,m}$, are small enough in magnitude in comparison with the elements of the channel transform matrix $\mathbf{H}_{n,m}$. Thus,

$$\begin{aligned}
\hat{\mathbf{x}}_{n,m} &\approx \mathbf{x}_{n,m} - (\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m})^{-1}[\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m} - \mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m}(\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m} + \mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{E}_{n,m})^{-1}\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m}]\mathbf{x}_{n,m} \\
&\quad + [\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m} + \mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{E}_{n,m}]^{-1}\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{w}_{n,m} \\
&= \mathbf{x}_{n,m} - (\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m} + \mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{E}_{n,m})^{-1}\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{E}_{n,m}\mathbf{x}_{n,m} + (\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m} + \mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{E}_{n,m})^{-1}\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{w}_{n,m} \\
&= \mathbf{x}_{n,m} + [\mathbf{H}^{\mathrm{H}}_{n,m}(\mathbf{H}_{n,m} + \mathbf{E}_{n,m})]^{-1}\mathbf{H}^{\mathrm{H}}_{n,m}(-\mathbf{E}_{n,m}\mathbf{x}_{n,m} + \mathbf{w}_{n,m}) \\
&\approx \mathbf{x}_{n,m} + (\mathbf{H}^{\mathrm{H}}_{n,m}\mathbf{H}_{n,m})^{-1}\mathbf{H}^{\mathrm{H}}_{n,m}(-\mathbf{E}_{n,m}\mathbf{x}_{n,m} + \mathbf{w}_{n,m}) \\
&= \mathbf{x}_{n,m} + \mathbf{H}^{+}_{n,m}(-\mathbf{E}_{n,m}\mathbf{x}_{n,m} + \mathbf{w}_{n,m}) = \mathbf{x}_{n,m} + \boldsymbol{\xi}_{n,m}
\end{aligned} \quad , \qquad \text{(D.2)}$$

where $\boldsymbol{\xi}_{n,m}$ represents the detection error term, which has zero mean, $\mathrm{E}[\boldsymbol{\xi}_{n,m}] = \mathbf{0}$, and the covariance matrix

$$\mathrm{E}[\boldsymbol{\xi}_{n,m}\boldsymbol{\xi}_{n,m}^{\mathrm{H}}] = \mathrm{E}[\mathbf{H}_{n,m}^{+}(-\mathbf{E}_{n,m}\mathbf{x}_{n,m}+\mathbf{w}_{n,m})(-\mathbf{x}_{n,m}^{\mathrm{H}}\mathbf{E}_{n,m}^{\mathrm{H}}+\mathbf{w}_{n,m}^{\mathrm{H}})\mathbf{H}_{n,m}^{+\mathrm{H}}]$$
$$= \mathbf{H}_{n,m}^{+}\mathrm{E}[\mathbf{E}_{n,m}\mathbf{x}_{n,m}\mathbf{x}_{n,m}^{\mathrm{H}}\mathbf{E}_{n,m}^{\mathrm{H}}+\mathbf{w}_{n,m}\mathbf{w}_{n,m}^{\mathrm{H}}]\mathbf{H}_{n,m}^{+\mathrm{H}} \quad . \tag{D.3}$$

Expectation in the middle term of (D.3) is taken with respect to $\mathbf{E}_{n,m}\mathbf{x}_{n,m}\mathbf{x}_{n,m}^{\mathrm{H}}\mathbf{E}_{n,m}^{\mathrm{H}}$ and $\mathbf{w}_{n,m}\mathbf{w}_{n,m}^{\mathrm{H}}$, which are associated with the given CFR realisation $\mathbf{H}_{n,m}$ at the *n*th subcarrier inside the *m*th block. $\mathrm{E}\left[\mathbf{w}_{n,m}^{\mathrm{H}}\mathbf{E}_{n,m}\mathbf{x}_{n,m}\right] = \mathrm{E}\left[\mathbf{x}_{n,m}^{\mathrm{H}}\mathbf{E}_{n,m}^{\mathrm{H}}\mathbf{w}_{n,m}\right] = 0$ in (D.3) due to the mutual independence of $\mathbf{w}_{n,m}$ and $\mathbf{E}_{n,m}\mathbf{x}_{n,m}$, and $\mathrm{E}[\mathbf{w}_{n,m}\mathbf{w}_{n,m}^{\mathrm{H}}] = \sigma_{\mathrm{w}}^{2}\mathbf{I}_{N_{\mathrm{rx}}\times N_{\mathrm{rx}}}$ as $\mathbf{w}_{n,m}$ is made up of IID CGRVs with variance σ_{w}^{2}. To describe the statistics of $\mathbf{E}_{n,m}\mathbf{x}_{n,m}\mathbf{x}_{n,m}^{\mathrm{H}}\mathbf{E}_{n,m}^{\mathrm{H}}$ we let

$$\mathbf{u}_{n,m} = \mathbf{E}_{n,m}\mathbf{x}_{n,m} = \sum_{i=0}^{N_{\mathrm{tx}}-1}(\mathbf{E}_{n,m})_{i}\,x_{n,m}(i) = \sigma_{\mathrm{d}}\sum_{i=0}^{N_{\mathrm{tx}}-1}(\mathbf{E}_{n,m})_{i}\,\mathrm{e}^{\mathrm{j}\arg[x_{n,m}(i)]} \quad , \tag{D.4}$$

where $(.)_{i}$ denotes *i*th column of a matrix, $x_{n,m}(i)$ are the elements of the vector $\mathbf{x}_{n,m}$, and $\mathrm{j} = \sqrt{-1}$. From (D.4) one can see that $\mathbf{u}_{n,m}$ represents a scaled sum of IID CGRVs with randomly shifted phases ($\mathrm{E}\left[(\mathbf{E}_{n,m})_{i}\right] = \mathbf{0}$, $\mathrm{E}\left[(\mathbf{E}_{n,m})_{i}(\mathbf{E}_{n,m})_{i}^{\mathrm{H}}\right] = \sigma_{\varepsilon n}^{2}\mathbf{I}_{N_{\mathrm{rx}}\times N_{\mathrm{rx}}}$ and $\mathrm{E}\left[(\mathbf{E}_{n,m})_{i}(\mathbf{E}_{n,m})_{j}^{\mathrm{H}}\right] = \mathbf{0}_{N_{\mathrm{rx}}\times N_{\mathrm{rx}}}$ if $i \neq j$ for $\forall i, j \in [0, N_{\mathrm{tx}}-1]$). Hence it follows that $\mathrm{E}\left[\mathbf{u}_{n,m}\right] = \mathbf{0}$ and $\mathrm{E}\left[\mathbf{u}_{n,m}\mathbf{u}_{n,m}^{\mathrm{H}}\right] = N_{\mathrm{tx}}\sigma_{\mathrm{d}}^{2}\sigma_{\varepsilon n}^{2}\mathbf{I}_{N_{\mathrm{rx}}\times N_{\mathrm{rx}}}$, where $\sigma_{\varepsilon n}^{2}$ is interpreted as the channel estimation MSE observed on the *n*th subcarrier. Note here that despite the use of distinct pilot subcarrier sets, $\sigma_{\varepsilon n}^{2}$ is approximately the same for different SLs (i.e. we assume it approximately constant for $\forall i \in [0, N_{\mathrm{tx}}-1]$) when the pilot pattern is independent, with the cyclic shift of pilot symbol positions associated with different Tx antennas (refer to the definition in Section 3.1 for details). This is because the difference in the pilot position for the distinct Tx antennas is no more than $N_{\mathrm{tx}}-1$ subcarriers where N_{tx} typically does not exceed 8-10, as shown in Section 4.1. Hence the channel estimator structures are very similar and are characterised by nearly the same error variance distribution across subcarriers (especially in the systems with larger number of subcarriers in the band). In the case of the equispaced arrangement of pilot subcarriers, $\sigma_{\varepsilon n}^{2}$ is indeed the same for all SLs as the estimators are identical (refer to Subsections 3.2.4, 3.3.5 and 3.4.7).

Using the identities deduced in the preceding paragraph, it is straightforward to show that MSE at the detector's output, corresponding to the *n*th subcarrier, is equal to

$$\sigma_{\xi n}^{2} = N_{\mathrm{tx}}^{-1}\mathrm{tr}(\mathrm{E}[\boldsymbol{\xi}_{n,m}\boldsymbol{\xi}_{n,m}^{\mathrm{H}}])$$
$$= N_{\mathrm{tx}}^{-1}\mathrm{tr}[\mathbf{H}_{n,m}^{+\mathrm{H}}\mathbf{H}_{n,m}^{+}\,\mathrm{E}(\mathbf{u}_{n,m}\mathbf{u}_{n,m}^{\mathrm{H}}+\mathbf{w}_{n,m}\mathbf{w}_{n,m}^{\mathrm{H}})]$$
$$= N_{\mathrm{tx}}^{-1}(N_{\mathrm{tx}}\sigma_{\mathrm{d}}^{2}\sigma_{\varepsilon n}^{2}+\sigma_{\mathrm{w}}^{2})\,\mathrm{tr}[\mathbf{H}_{n,m}(\mathbf{H}_{n,m}^{\mathrm{H}}\mathbf{H}_{n,m})^{-2}\mathbf{H}_{n,m}^{\mathrm{H}}]$$
$$= (\sigma_{\mathrm{d}}^{2}\sigma_{\varepsilon n}^{2}+N_{\mathrm{tx}}^{-1}\sigma_{\mathrm{w}}^{2})\,\mathrm{tr}[(\mathbf{H}_{n,m}^{\mathrm{H}}\mathbf{H}_{n,m})^{-1}] \quad . \tag{D.5}$$

D.2: Characterisation of detection NMSE as an optimisation function (CMMSE case)

The NMSE function (4.19),

$$\gamma = c_{\mathrm{H}}(P N^{-1}\kappa + b + P N_{\mathrm{tx}} SNR \,\mathrm{tr}[\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}\mathbf{S}\mathbf{U}_{\underline{\mathbf{hh}}}\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}}(P\,SNR\,\kappa/(P N^{-1}\kappa + b)\,\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I})^{-1}]), \tag{D.6}$$

is monotonic in the PDR argument κ. Hence it is differentiable with respect to κ. Here we determine the first and second derivatives of γ to characterise behaviour of the function.

Using the definitions of the matrix trace and inverse derivatives from [138], $\partial\,\mathrm{tr}(\mathbf{A})/\partial u = \mathrm{tr}(\partial\mathbf{A}/\partial u)$ and $\partial(\mathbf{A}^{-1})/\partial u = -\mathbf{A}^{-1}\,\partial\mathbf{A}/\partial u\,\mathbf{A}^{-1}$, one can obtain

$$\begin{aligned}
\frac{\partial\gamma}{\partial\kappa} &= c_{\mathrm{H}}\left(P N^{-1} - P N_{\mathrm{tx}} SNR\,\mathrm{tr}\left[\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}\mathbf{S}\mathbf{U}_{\underline{\mathbf{hh}}}\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}}\left(\frac{P\,SNR\,\kappa}{(P N^{-1}\kappa + b)}\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I}\right)^{-2}\frac{\partial}{\partial\kappa}\left(\frac{P\,SNR\,\kappa}{(P N^{-1}\kappa + b)}\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I}\right)\right]\right) \\
&= c_{\mathrm{H}}\left(P N^{-1} - P^2 N_{\mathrm{tx}} SNR^2\left(\frac{1}{P N^{-1}\kappa + b} - \frac{P N^{-1}\kappa}{(P N^{-1}\kappa + b)^2}\right)\mathrm{tr}\left[\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}\mathbf{S}\mathbf{U}_{\underline{\mathbf{hh}}}\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}}^2\left(\frac{P\,SNR\,\kappa}{(P N^{-1}\kappa + b)}\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} + \mathbf{I}\right)^{-2}\right]\right) \\
&= c_{\mathrm{H}}\left(P N^{-1} - P^2 N_{\mathrm{tx}} SNR^2 b\,\mathrm{tr}\left[\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}\mathbf{S}\mathbf{U}_{\underline{\mathbf{hh}}}\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}}^2\left(P(SNR\,\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} + N^{-1}\mathbf{I})\,\kappa + b\mathbf{I}\right)^{-2}\right]\right) \\
&= c_{\mathrm{H}} P N^{-1}\left(1 - P N_{\mathrm{tx}} N\,SNR^2 b\,\mathrm{tr}\left[\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}\mathbf{S}\mathbf{U}_{\underline{\mathbf{hh}}}\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}}^2\left(P(SNR\,\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} + N^{-1}\mathbf{I})\,\kappa + b\mathbf{I}\right)^{-2}\right]\right),
\end{aligned} \tag{D.7}$$

Subsequent differentiation yields

$$\begin{aligned}
\frac{\partial^2\gamma}{\partial\kappa^2} &= c_{\mathrm{H}} P^2 N_{\mathrm{tx}} SNR^2 b\,\mathrm{tr}\left[\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}\mathbf{S}\mathbf{U}_{\underline{\mathbf{hh}}}\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}}^2\left(P(SNR\,\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} + N^{-1}\mathbf{I})\,\kappa + b\mathbf{I}\right)^{-4}\frac{\partial}{\partial\kappa}\left(P(SNR\,\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} + N^{-1}\mathbf{I})\,\kappa + b\mathbf{I}\right)^2\right] \\
&= 2 c_{\mathrm{H}} P^3 N_{\mathrm{tx}} SNR^2 b\,\mathrm{tr}\left[\mathbf{S}^{1/2}\mathbf{U}_{\underline{\mathbf{hh}}}\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}}^2(SNR\,\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} + N^{-1}\mathbf{I})\left(P(SNR\,\boldsymbol{\Lambda}_{\underline{\mathbf{hh}}} + N^{-1}\mathbf{I})\,\kappa + b\mathbf{I}\right)^{-3}\mathbf{U}_{\underline{\mathbf{hh}}}^{\mathrm{H}}\mathbf{S}^{1/2}\right].
\end{aligned} \tag{D.8}$$

As the trace of the positive definite matrix is a positive number, $\partial^2\gamma/\partial\kappa^2 > 0$, and hence γ represents a convex function in κ.

D.3: Functional dependence of effective channel model order on predicted SNR (CMMSE case)

According to (4.31), the effective order of the channel model,

$$L_{\mathrm{eff}} = \bar{\rho}\left[\left(\sum_{l=0}^{L-1}\frac{\lambda_l}{1+\lambda_l\bar{\rho}}\right)^{-1} - 1\right]^{-1}, \tag{D.9}$$

represents a monotonic function of $\widehat{SNR}_{\mathrm{p}} > 0$. To ascertain the functional behaviour, we differentiate the inverse of L_{eff} with respect to $\hat{\rho} = P\,\widehat{SNR}_{\mathrm{p}}$, yielding

$$\begin{aligned}\frac{\partial L_{\mathrm{eff}}^{-1}}{\partial \hat{\rho}} &= -\frac{1}{\hat{\rho}^2}\left[\left(\sum_{l=0}^{L-1}\frac{\lambda_l}{1+\lambda_l\hat{\rho}}\right)^{-1}-1\right]+\frac{1}{\hat{\rho}}\left[\sum_{l=0}^{L-1}\left(\frac{\lambda_l}{1+\lambda_l\hat{\rho}}\right)^2\right]\left[\sum_{l=0}^{L-1}\frac{\lambda_l}{1+\lambda_l\hat{\rho}}\right]^{-2} \\ &= \frac{-\left(1-\sum_{l=0}^{L-1}\frac{\lambda_l}{1+\lambda_l\hat{\rho}}\right)\left(\sum_{l=0}^{L-1}\frac{\lambda_l}{1+\lambda_l\hat{\rho}}\right)+\hat{\rho}\sum_{l=0}^{L-1}\left(\frac{\lambda_l}{1+\lambda_l\hat{\rho}}\right)^2}{\hat{\rho}^2\left(\sum_{l=0}^{L-1}\frac{\lambda_l}{1+\lambda_l\hat{\rho}}\right)^2} \quad . \end{aligned} \tag{D.10}$$

Substituting $\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}] = \sum_{l=0}^{L-1}\frac{\lambda_l}{1+\lambda_l\hat{\rho}}$ to make the notation more compact, one can re-express $\partial L_{\mathrm{eff}}^{-1}/\partial\hat{\rho}$ as

$$\begin{aligned}\frac{\partial L_{\mathrm{eff}}^{-1}}{\partial \hat{\rho}} &= \frac{-\left(1-\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{I}+\hat{\rho}\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}})^{-1}]\right)\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{I}+\hat{\rho}\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}})^{-1}]+\hat{\rho}\,\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}^2(\mathbf{I}+\hat{\rho}\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}})^{-2}]}{\hat{\rho}^2\left(\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{I}+\hat{\rho}\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}})^{-1}]\right)^2} \\ &= \frac{-\left(1-\hat{\rho}^{-1}\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]\right)\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]+\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}^2(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-2}]}{\hat{\rho}^3\left(\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{I}+\hat{\rho}\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}})^{-1}]\right)^2} \\ &= \frac{-\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]+\hat{\rho}^{-1}\left(\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]\right)^2+\mathrm{tr}[\mathbf{I}-(2\hat{\rho}^{-1}\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-2}\mathbf{I})(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-2}]}{\hat{\rho}^3\left(\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{I}+\hat{\rho}\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}})^{-1}]\right)^2} \\ &= \frac{-L\hat{\rho}+\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]+\left(L-\hat{\rho}^{-1}\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]\right)^2+L\hat{\rho}-\mathrm{tr}[(2\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-2}]}{\hat{\rho}^4\left(\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{I}+\hat{\rho}\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}})^{-1}]\right)^2} \\ &= \frac{L^2+\hat{\rho}^{-2}[\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]^2-2L\hat{\rho}^{-1}\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]-\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-2}]}{\hat{\rho}^4\left(\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{I}+\hat{\rho}\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}})^{-1}]\right)^2} \\ &= \frac{L^2+\hat{\rho}^{-2}[\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]^2-\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}(2L\hat{\rho}^{-1}\mathbf{I}+\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1})]}{\hat{\rho}^4\left(\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{I}+\hat{\rho}\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}})^{-1}]\right)^2} \\ &= \frac{L^2+\hat{\rho}^{-2}[\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]^2-\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}((2L\hat{\rho}^{-1}+1)\mathbf{I}-\hat{\rho}^{-1}(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1})]}{\hat{\rho}^4\left(\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{I}+\hat{\rho}\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}})^{-1}]\right)^2} \\ &= \frac{L^2-(2L\hat{\rho}^{-1}+1)\,\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]+\hat{\rho}^{-1}\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-2}]+\hat{\rho}^{-2}[\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]^2}{\hat{\rho}^4\left(\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{I}+\hat{\rho}\,\mathbf{\Lambda}_{\underline{\mathbf{hh}}})^{-1}]\right)^2} \quad , \end{aligned} \tag{D.11}$$

where the identity $\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}] = L-\hat{\rho}^{-1}\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]$ [138] has been used repeatedly. For the high SNR regime ($\hat{\rho}\to\infty$), the term $L^2-(2L\hat{\rho}^{-1}+1)\,\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]$ represents the biggest contribution to the numerator value. Using the trace version of the Cauchy-Schwarz inequality [131], $\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]\,\mathrm{tr}[\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I}] \geq L^2$, one can determine the maximum of $L^2-(2L\hat{\rho}^{-1}+1)\,\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}]$ and hence an upper limit on $\partial L_{\mathrm{eff}}^{-1}/\partial\hat{\rho}$. As $\mathrm{tr}[(\mathbf{\Lambda}_{\underline{\mathbf{hh}}}+\hat{\rho}^{-1}\mathbf{I})^{-1}] \geq \hat{\rho}\,L^2/(\hat{\rho}+L)$, with the equality if and only if $\lambda_l = L^{-1}, \forall\, l\in[0,L-1]$, we can put down after substitution of $\mathbf{\Lambda}_{\underline{\mathbf{hh}}} = L^{-1}\mathbf{I}$ in (D.11) that

$$\frac{\partial L_{\text{eff}}^{-1}}{\partial \widehat{\rho}} \leq \frac{L^2 - \dfrac{L^2\widehat{\rho}(2L\widehat{\rho}^{-1}+1)}{(\widehat{\rho}+L)} + \dfrac{1}{\widehat{\rho}}\left[\dfrac{L}{(\widehat{\rho}^{-1}+L^{-1})^2}\right] + \dfrac{1}{\widehat{\rho}^2}\left[\dfrac{L^2\widehat{\rho}}{\widehat{\rho}+L}\right]^2}{\widehat{\rho}^4\left(\widehat{\rho}\,L/(\widehat{\rho}+L)\right)^2}. \tag{D.12}$$

Performing trivial reductions on the right side of (D.12), it is established that

$$\frac{\partial L_{\text{eff}}^{-1}}{\partial \widehat{\rho}} \leq \frac{L^4 - L^3(\widehat{\rho}+L) + L^3\widehat{\rho}}{L^2\widehat{\rho}^6} = 0, \tag{D.13}$$

from where it becomes clear that L_{eff} is a non-decreasing function of $\widehat{SNR}_{\text{p}}$.

D.4: Characterisation of detection NMSE as optimisation function (CLS-MMSE case)

Given the detection NMSE function (4.35),

$$\gamma = c_{\text{H}}\left[\frac{P}{N}\kappa + b + N_{\text{tx}}\left(\frac{P}{N}+\frac{b}{\kappa}\right)\sum_{l=0}^{L-1} c_l\left[1-\left(\frac{P\kappa+Nb}{a_l\kappa+Nb}\right)^{\Omega^{-1}}\right]\right], \tag{D.14}$$

one can express the first and the second-order derivatives as

$$\begin{aligned}
\frac{\partial \gamma}{\partial \kappa} &= c_{\text{H}}\left[\frac{P}{N} - \frac{N_{\text{tx}}b}{\kappa^2}\sum_{l=0}^{L-1} c_l\left[1-\left(\frac{P\kappa+Nb}{a_l\kappa+Nb}\right)^{\Omega^{-1}}\right]\right. \\
&\quad \left. - N_{\text{tx}}\left(\frac{P}{N}+\frac{b}{\kappa}\right)\sum_{l=0}^{L-1}\frac{c_l}{\Omega}\left(\frac{P\kappa+Nb}{a_l\kappa+Nb}\right)^{\Omega^{-1}-1}\left(\frac{P}{a_l\kappa+Nb} - \frac{a_l(P\kappa+Nb)}{(a_l\kappa+Nb)^2}\right)\right] \\
&= c_{\text{H}}\left[\frac{P}{N} - \frac{N_{\text{tx}}b}{\kappa^2}\sum_{l=0}^{L-1} c_l\right. \\
&\quad \left. + N_{\text{tx}}\sum_{l=0}^{L-1} c_l\left(\frac{P\kappa+Nb}{a_l\kappa+Nb}\right)^{\Omega^{-1}}\left(\frac{b}{\kappa^2} - \frac{(a_l\kappa+Nb)}{\Omega N\kappa}\frac{[(Pa_l - Pa_l)\kappa + PNb - Nba_l]}{(a_l\kappa+Nb)^2}\right)\right] \\
&= c_{\text{H}}\left[\frac{P}{N} - \frac{N_{\text{tx}}b}{s\kappa^2} + \frac{N_{\text{tx}}b}{\kappa}\sum_{l=0}^{L-1} c_l\left(\frac{P\kappa+Nb}{a_l\kappa+Nb}\right)^{\Omega^{-1}}\left(\frac{1}{\kappa} + \frac{a_l - P}{\Omega(a_l\kappa+Nb)}\right)\right],
\end{aligned} \tag{D.15}$$

$$\frac{\partial^2\gamma}{\partial\kappa^2} = c_{\mathrm{H}}\left[\frac{2N_{\mathrm{tx}}b}{s\kappa^3} + N_{\mathrm{tx}}b\sum_{l=0}^{L-1}c_l\left[\frac{1}{\Omega}\left(\frac{P\kappa+Nb}{a_l\kappa+Nb}\right)^{\Omega^{-1}-1}\left(\frac{P}{a_l\kappa+Nb}-\frac{a_l(P\kappa+Nb)}{(a_l\kappa+Nb)^2}\right)\left(\frac{1}{\kappa^2}+\frac{a_l-P}{\Omega\kappa(a_l\kappa+Nb)}\right)\right.\right.$$
$$\left.\left.-\left(\frac{P\kappa+Nb}{a_l\kappa+Nb}\right)^{\Omega^{-1}}\left(\frac{2}{\kappa^3}+\frac{(a_l-P)(2a_l\kappa+Nb)}{\Omega\kappa^2(a_l\kappa+Nb)^2}\right)\right]\right]$$
$$= c_{\mathrm{H}}\left[\frac{2N_{\mathrm{tx}}b}{s\kappa^3} - N_{\mathrm{tx}}b\sum_{l=0}^{L-1}c_l\left(\frac{P\kappa+Nb}{a_l\kappa+Nb}\right)^{\Omega^{-1}}\left[\frac{Nb(a_l-P)}{\Omega(P\kappa+Nb)(a_l\kappa+Nb)}\left(\frac{1}{\kappa^2}+\frac{a_l-P}{\Omega\kappa(a_l\kappa+Nb)}\right)\right.\right.$$
$$\left.\left.+\frac{2}{\kappa^3}+\frac{(a_l-P)(2a_l\kappa+Nb)}{\Omega\kappa^2(a_l\kappa+Nb)^2}\right]\right]$$
$$= c_{\mathrm{H}}\left[\frac{2N_{\mathrm{tx}}b}{\kappa^3}\sum_{l=0}^{L-1}c_l\left[1-\Gamma_l\left(\frac{P\kappa+Nb}{a_l\kappa+Nb}\right)^{\Omega^{-1}}\right]\right], \quad \text{(D.16)}$$

where

$$\Gamma_l = 1+\frac{1}{2\Omega}\left(\frac{Nb(a_l-P)}{(P\kappa+Nb)(a_l+Nb\kappa^{-1})}+\frac{\Omega^{-1}Nb(a_l-P)^2}{(P\kappa+Nb)(a_l+Nb\kappa^{-1})^2}\right.$$
$$\left.+\frac{a_l-P}{a_l+Nb\kappa^{-1}}+\frac{a_l(a_l-P)}{(a_l+Nb\kappa^{-1})^2}\right)$$
$$= 1+\frac{1}{2\Omega}\left(\frac{Nb(a_l-P)}{(P\kappa+Nb)(a_l+Nb\kappa^{-1})}+\frac{\Omega^{-1}Nb(a_l-P)^2}{(P\kappa+Nb)(a_l+Nb\kappa^{-1})^2}\right.$$
$$\left.+\frac{a_l-P}{a_l+Nb\kappa^{-1}}+\frac{(a_l-P)^2}{(a_l+Nb\kappa^{-1})^2}+\frac{P(a_l-P)}{(a_l+Nb\kappa^{-1})^2}\right)$$
$$= 1+\frac{1}{2\Omega}\left(\frac{(a_l-P)(P\kappa+2Nb)}{\kappa(P+Nb\kappa^{-1})(a_l+Nb\kappa^{-1})}\right.$$
$$\left.+\frac{(a_l-P)^2(\Omega^{-1}Nb+Nb+P\kappa)}{\kappa(P+Nb\kappa^{-1})(a_l+Nb\kappa^{-1})^2}+\frac{P(a_l-P)}{(a_l+Nb\kappa^{-1})^2}\right)$$
$$= 1+\frac{1}{2\Omega}\left[\frac{(a_l-P)}{(a_l+Nb\kappa^{-1})}\left(\frac{P+2Nb\kappa^{-1}}{P+Nb\kappa^{-1}}+\frac{P}{a_l+Nb\kappa^{-1}}\right)\right.$$
$$\left.+\frac{(1+\Omega^{-1})(a_l-P)^2[P(1+\Omega^{-1})^{-1}+Nb\kappa^{-1}]}{(P+Nb\kappa^{-1})(a_l+Nb\kappa^{-1})^2}\right], \quad \text{(D.17)}$$

and it is evident that

$$\Gamma_l < 1+\frac{1}{\Omega}\left(\frac{a_l-P}{a_l+Nb\kappa^{-1}}+\frac{(1+\Omega^{-1})(a_l-P)^2}{2(P+Nb\kappa^{-1})(a_l+Nb\kappa^{-1})}\right) \quad \text{(D.18)}$$

since $a_l \geq P$ and $\Omega \geq 1$.

To prove that $\partial^2\gamma/\partial\kappa^2 > 0$, implying convexity of γ, we use the asymptotically converging series expansion [103]:

$$\left(\frac{a_l\kappa+N\,b}{P\,\kappa+N\,b}\right)^{\Omega^{-1}}=1+\sum_{m=1}^{\infty}\frac{1}{m!}\left[\frac{1}{\Omega}\ln\left(\frac{a_l\kappa+N\,b}{P\,\kappa+N\,b}\right)\right]^m, \tag{D.19}$$

where

$$\ln\left(\frac{a_l\kappa+N\,b}{P\,\kappa+N\,b}\right)=\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{a_l-P}{P+N\,b\,\kappa^{-1}}\right)^n. \tag{D.20}$$

Limiting expansion (D.19) to only the first two terms, the first of which is restricted to the first two terms of (D.20) and the second is constrained by just the starting term, we can formulate the following inequality:

$$\begin{aligned}\left(\frac{a_l\kappa+N\,b}{P\,\kappa+N\,b}\right)^{\Omega^{-1}}&>1+\frac{1}{\Omega}\left(\frac{a_l-P}{P+N\,b\,\kappa^{-1}}+\frac{(a_l-P)^2}{2(P+N\,b\,\kappa^{-1})^2}+\frac{(a_l-P)^2}{2\Omega(P+N\,b\,\kappa^{-1})^2}\right)\\&=1+\frac{1}{\Omega}\left(\frac{a_l-P}{P+N\,b\,\kappa^{-1}}+\frac{(1+\Omega^{-1})(a_l-P)^2}{2(P+N\,b\,\kappa^{-1})^2}\right)>\Gamma_l\end{aligned}. \tag{D.21}$$

Hence $\Gamma_l\left(\frac{P\,\kappa+N\,b}{a_l\kappa+N\,b}\right)^{\Omega^{-1}}<1$ in (D.16) and $\partial^2\gamma/\partial\kappa^2>0$, meaning that the function (D.14) is convex in the interval $\kappa\in(0,\infty)$.

D.5: Derivation of upper bound on the optimal PDR (CLS-MMSE case)

It has been established in Appendix D.4 that the cost function (4.35) has a convex shape in the interval $\kappa\in(0,\infty)$. Hence one of the roots of (4.50), matching the local minimum in this interval, is real positive. Denote it κ_1, while the other two roots – κ_2 and κ_3. Referring to the relationship between the roots and the coefficients of the third-order polynomial [103],

$$\kappa_1\kappa_2\kappa_3=-C>0, \tag{D.22}$$

we can notice that the identity (D.22) is satisfied if and only if κ_2 and κ_3 are either both real negative or complex-conjugated. As shown in the following steps, the complexity property of κ_2 and κ_3 directly affects choice of the solving method.

Substitution of $\kappa=x-A/3$ in (4.50) leads to the depressed cubic

$$F_3(\kappa)=x^3+U\,x+V, \tag{D.23}$$

where

$$U=B-A^2/3, \tag{D.24}$$

$$V=2A^3/27-AB/3+C, \tag{D.25}$$

and that is characterised by the discriminant

$$D_3 = U^3/27 + V^2/4 . \tag{D.26}$$

The roots of the depressed cubic can be found using the general Cardano's formula [103]:

$$x_1 = \left(-V/2 + \sqrt{D_3}\right)^{1/3} + \left(-V/2 - \sqrt{D_3}\right)^{1/3}, \tag{D.27}$$

$$x_{2,3} = 1/2\left[-x_1 \pm \left(\left(-V/2 + \sqrt{D_3}\right)^{1/3} - \left(-V/2 - \sqrt{D_3}\right)^{1/3}\right)\sqrt{3}\mathrm{j}\right]. \tag{D.28}$$

However, Cardano's approach is not convenient as it involves complex number computations, e.g., when $D_3 < 0$. Special methods can be applied, depending on the sign of D_3.

To proceed with D_3 analysis, we first analyse the coefficients A (4.51), B (4.52) and C (4.53), as well as the term $AB/3$, contributing to the coefficient V (D.25) of the depressed cubic.

For A, B and C, the following relations hold:

$$0 < A << 1, \tag{D.29}$$

$$C < \frac{L^2 N_{\mathrm{tx}} N b^2}{P^3 SNR\,\Omega}\left(\bar{\varphi} - \frac{P}{Ls}\right) < 0, \tag{D.30}$$

due to $a > \Omega P L^{-1} N\, SNR$ (4.47) and $s \le P/L$ (following from the derivations in Appendix C.5), and

$$B < \frac{L N_{\mathrm{tx}} N b}{P^2}\left[\bar{\varphi}\left(1 + \frac{1}{\Omega}\right) - \frac{P}{Ls}\right] < 0, \tag{D.31}$$

as $\bar{\varphi} < \dfrac{\Omega}{\Omega + 1}$, which is proved in the sequel.

By substituting (4.47) into (4.49), we can put down that

$$\bar{\varphi} = \left(1 + \frac{\Omega P L^{-1} N\, SNR}{P + N b \kappa_{\mathrm{eff}}^{-1}}\right)^{-\Omega^{-1}} . \tag{D.32}$$

Taking into account that according to (4.40) and (4.12),

$$\kappa_{\mathrm{eff}} \ge P^{-1}\sqrt{L N_{\mathrm{tx}} N b}, \tag{D.33}$$

$$b/N_{\mathrm{tx}} < 1, \tag{D.34}$$

and that according to Subsection 2.2.4, the block-transmitting systems are typically designed for

$$L/N \le 1/4, \tag{D.35}$$

$\bar{\varphi}$ (D.32) is limited by the inequality

$$\bar{\varphi} = \left(1 + \frac{\Omega\, SNR}{LN^{-1} + \sqrt{b N_{\mathrm{tx}}^{-1} LN^{-1}}}\right)^{-\Omega^{-1}} < \left(1 + \frac{4}{3} SNR\,\Omega\right)^{-\Omega^{-1}} . \tag{D.36}$$

Using the series expansions [103]

$$\ln\left(1 + \frac{4}{3} SNR\,\Omega\right) = \sum_{m=1}^{\infty} \frac{1}{m}\left(\frac{4/3\, SNR\,\Omega}{1 + 4/3\, SNR\,\Omega}\right)^m , \tag{D.37}$$

$$\mathrm{e}^{x} = 1 + \sum_{n=1}^{\infty} \frac{x^{n}}{n!}, \tag{D.38}$$

constrained to the first two terms, (D.36) can be further reduced to

$$\begin{aligned}
\bar{\varphi} &< \exp\left[-\Omega^{-1}\ln\left(1+4/3\,SNR\,\Omega\right)\right] \\
&< \exp\left[-\Omega^{-1}\left(\frac{4/3\,SNR\,\Omega}{1+4/3\,SNR\,\Omega}+\frac{(4/3\,SNR\,\Omega)^{2}}{2(1+4/3\,SNR\,\Omega)^{2}}\right)\right] \\
&= \exp\left[\frac{4/3\,SNR+8/3\,SNR^{2}\Omega}{(1+4/3\,SNR\,\Omega)^{2}}\right]^{-1} \\
&< \left(1+\frac{4/3\,SNR\,(1+2SNR\,\Omega)}{(1+4/3\,SNR\,\Omega)^{2}}\right)^{-1}
\end{aligned} . \tag{D.39}$$

Calculation of the difference inequality

$$\begin{aligned}
\frac{1}{\bar{\varphi}} - \left(1+\frac{1}{\Omega}\right) &= \frac{4/3\,SNR\,(1+2SNR\,\Omega)}{(1+4/3\,SNR\,\Omega)^{2}} - \frac{1}{\Omega} \\
&= \frac{8/9\,SNR^{2}\Omega^{2} - 4/3\,SNR\,\Omega - 1}{\Omega\,(1+4/3\,SNR\,\Omega)^{2}} > 0
\end{aligned} \tag{D.40}$$

allows one to establish that the inequality $\bar{\varphi} < \left(1+\Omega^{-1}\right)^{-1}$ holds for $SNR\,\Omega > 3/4\,(1+\sqrt{3}) \approx 2.05$. This condition is typically satisfied for all $SNR > 2$, irrespective of the Ω value. Hence it follows that $B < 0$.

With regard to $AB/3$, it can be established that the following inequality holds:

$$\frac{AB}{3} = \frac{N_{\mathrm{tx}}N^{2}b^{2}}{Pa}\left(\frac{N\,SNR\,\bar{\varphi}}{a}+\frac{L\bar{\varphi}}{P}-\frac{1}{s}\right) < \frac{L^{2}N_{\mathrm{tx}}Nb^{2}}{3P^{3}SNR\,\Omega}\left[\bar{\varphi}\left(1+\frac{1}{\Omega}\right)-\frac{P}{Ls}\right] < 0\,. \tag{D.41}$$

Comparing $AB/3$ and C (4.53), one can notice that $|C| > |AB/3|$. Hence as $2A^{3}/27 \to 0$,

$$V = -\left(AB/3 - C\right) + 2A^{3}/27 > -\left(AB/3 - C\right), \tag{D.42}$$

where $AB/3 - C - 2A^{3}/27 > 0$, so that

$$\begin{aligned}
V^{2} < \left(AB/3 - C\right)^{2} &= \left[\frac{N_{\mathrm{tx}}N^{2}b^{2}}{3Pa}\left(\frac{N\,SNR\,\bar{\varphi}}{a}-\frac{2L\bar{\varphi}}{P}+\frac{2}{s}\right)\right]^{2} \\
&< \left[\frac{2LN_{\mathrm{tx}}Nb^{2}}{3P^{2}SNR\,\Omega}\left(\frac{L\bar{\varphi}}{2P\Omega}-\frac{L\bar{\varphi}}{P}+\frac{1}{s}\right)\right]^{2} = \frac{4L^{4}N_{\mathrm{tx}}^{2}N^{2}b^{4}}{9P^{6}SNR^{2}\Omega^{2}}\left[\frac{P}{Ls}-\bar{\varphi}\left(1-\frac{1}{2\Omega}\right)\right]^{2}
\end{aligned} . \tag{D.43}$$

Using the inequalities (D.34), (D.35) and $SNR > 2$, we can further constrain V^{2} to

$$V^{2} < \frac{L^{3}N_{\mathrm{tx}}^{3}N^{3}b^{3}}{36P^{6}\Omega^{2}}\left[\frac{P}{Ls}-\bar{\varphi}\left(1-\frac{1}{2\Omega}\right)\right]^{2}. \tag{D.44}$$

In analogy with (D.42), the coefficient U can be expressed as

$$U = B - A^{2}/3 < B\,, \tag{D.45}$$

where $-B + A^{2}/3 > 0$ as $A^{2}/3 \to 0$, so that

$$U^3 < B^3 = -\frac{L^3 N_{tx}^3 N^3 b^3}{P^6}\left[\frac{P}{Ls} - \bar{\varphi}\left(1+\frac{1}{\Omega}\right)\right]^3. \tag{D.46}$$

Substitution of (D.44) and (D.46) into (D.26) yields

$$D_3 < \frac{L^3 N_{tx}^3 N^3 b^3}{9P^6} d_3, \tag{D.47}$$

with

$$d_3 = -\frac{1}{3}\left[\frac{P}{Ls} - \bar{\varphi}\left(1+\frac{1}{\Omega}\right)\right]^3 + \frac{1}{16\Omega^2}\left[\frac{P}{Ls} - \bar{\varphi}\left(1-\frac{1}{2\Omega}\right)\right]^2, \tag{D.48}$$

and

$$\bar{\varphi} = \left(1+W\,\Omega\right)^{-\Omega^{-1}}, \tag{D.49}$$

where according to (D.36) $W > 4/3\,SNR > 8/3$.

From (D.48), one can see that d_3 is a monotonic function of the two parameters, $\Omega \in [1,\infty)$ and $W \in (8/3,\infty)$. Investigation of the properties of d_3 using the classical analytical procedures is too sophisticated. Therefore the numerical evaluation of d_3 has been performed for the equispaced pilot pattern ($s = P/L$). One can clearly see the asymptotic trends of the two-dimensional d_3 graph in Fig.D.1. Over the entire parameter plane the function takes negative values. $D_3 < 0$ implies narrowing of the cubic solution space to only three distinct real roots. Although the graph corresponds to the equispaced pilot arrangement, for other pilot patterns the same trends hold in the function's behaviour, as with $s < P/L$, the cubic term in the sum (D.48) even more prevails over the quadratic term than for $s = P/L$.

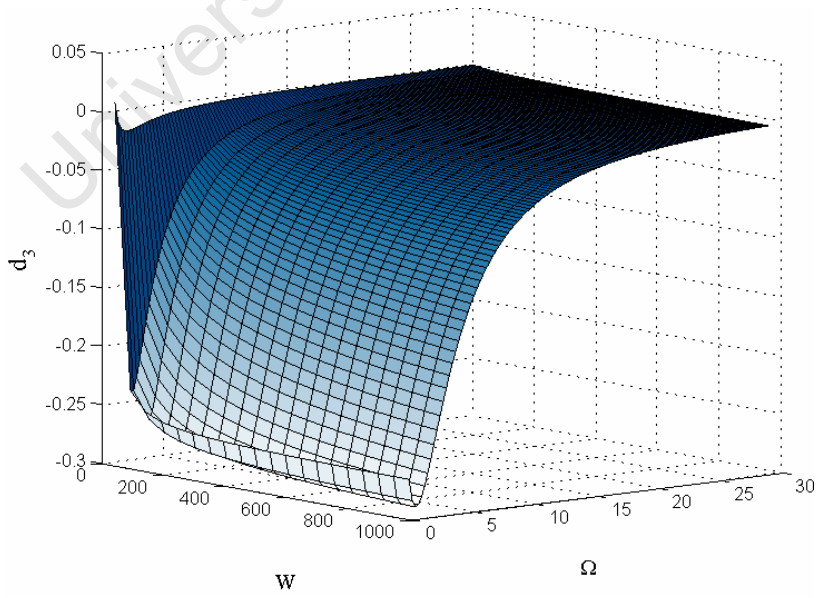


Fig.D.1. d_3 as a function of Ω and W

For the case of $D_3 < 0$, a simple trigonometric root finding method has been developed [142] instead of the general solution (D.27)-(D.28), involving complex plane computations. We recommend this method to calculate $\widehat{\kappa}_{\text{opt}}$. Define

$$\vartheta = \arccos\left(-\frac{3}{2}\sqrt{-\frac{3V^2}{U^3}} \right). \tag{D.50}$$

Then the three roots of (D.23) are given by

$$x_1 = \frac{2}{3}\sqrt{-\frac{U}{3}}\cos\left(\frac{\vartheta}{3}\right), \tag{D.51}$$

$$x_2 = \frac{2}{3}\sqrt{-\frac{U}{3}}\cos\left(\frac{\vartheta+2\pi}{3}\right), \tag{D.52}$$

$$x_3 = \frac{2}{3}\sqrt{-\frac{U}{3}}\cos\left(\frac{\vartheta+4\pi}{3}\right). \tag{D.53}$$

$0 \le \vartheta \le \pi$ and hence $\vartheta/3 \in [0, \pi/3]$ implies that $x_1 \ge 0$, whereas $2\pi/3 \le (\vartheta+2\pi)/3 \le \pi$ and $4\pi/3 \le (\vartheta+4\pi)/3 \le 5\pi/3$ lead to $x_2 < 0$ and $x_3 \le x_1$ as $\cos((\vartheta+4\pi)/3) \le \cos(5\pi/3) = \cos(\pi/3)$. Hence the solution is uniquely selected from the largest root x_1:

$$\widehat{\kappa}_{\text{opt}} = x_1 - A/3 = \frac{2}{3}\sqrt{A^2 - 3B}\cos\left[\frac{1}{3}\arccos\left(\frac{9AB - 27C - 2A^3}{2\sqrt{(A^2-3B)^3}}\right)\right] - \frac{A}{3}. \tag{D.54}$$